

Formative turbulent months ahead

The British research community appears to be awakening to the radicalism of its government's new plans. Prudence requires that the outlines should be accepted, but the details cry out for argument.

IT is curious that a few tentative comments (*Nature* 328,279; 1987), some of them ironic, on the British government's new plans for universities and the conduct of publicly sponsored research should have drawn some of the first responses from British readers in three years to a largely consistent and mostly despondent commentary on the condition of British science. It would have been helpful if those who now fear that *Nature* accepts the government's evaluation of the utility of the British research enterprise ("With friends like this...") had been moved to chip in at an earlier stage, perhaps when the British government was accused of having arranged, by "parsimony, indolence and indifference", that "the goose that might have laid those golden eggs [of research-led prosperity] will not be around much longer" (*Nature* 310, 261; 1984). Then, at least, it might have been possible more successfully to influence the government's train of thought leading to the announcement last month of what it intends to do.

The issues raised by D.A.W. Grant and Bruce Charlton on page 754 of this issue are nonetheless important and timely. They may be summarized as follows:

Is British science bankrupt? No. There is still a wealth of important basic research under way in British laboratories, in universities and public research institutes. By coincidence, this issue of *Nature* is a good illustration that the kind of work that gave the British a reputation for innovation in research until the early 1970s still survives. The identification of the slime mould morphogen described on pages 822–824 is an example of how people are still able to concentrate resources on the execution of a taxing project; the neat way in which experimental plants appear to be protected against viral infection (pp. 810–813) shows that even the much-harassed British agricultural research institutes have life left in them. It is even possible that, in the years ahead, adversity may bring out the best in the research teams still left intact. The worry is not so much with the quality of what is now being done, but with the future. For want of sufficient renewal over 15 years, the research community is ageing. The depth and variety of its pattern of work, already constrained by the lack of funds, will be further restricted by the reorganizations now on the cards. There are good reasons to fear that the permanent loss of able people is potentially another undermining influence. The flight of able young people into fields other than science, made possible and even necessary by the British educational system, is a greater if more distant worry. Bankruptcy tomorrow is the threat.

Who is to blame? The government, especially in Sir Keith (now Lord) Joseph's time, must shoulder much of the responsibility. "Parsimony, indolence and indifference" have been much in evidence. Many enforced reorganizations, as of the agricultural research institutes, have saved very little money but have undermined morale. But it is also true that the research community's natural leadership has been too passive, too silent and, when it has spoken out, too unconstructive. For much of the past seven years, the government would probably have backed any reasoned plan by the research community for spending the same amount of money differently. Now, under different management, it plans to do the job itself. There will be a new

advisory committee, a pecking order among universities and a concentration of research funds on projects calculated to open up fields with scientific interest that will nevertheless yield commercial benefits. At least the chairmen of the Advisory Board for the Research Councils (ABRC) and of the Science and Engineering Research Council apparently agree. Whether they will get the funds with which to launch their schemes is an open question.

Are the new arrangements wise? Who can say? The British have a penchant for launching major reorganizations of civil science (as in 1963, 1971, 1978 — abortively — and now). This time, the reorganization will go deeper, touching higher education radically (but probably leaving the research councils as they are). The government's plan goes some way towards meeting the demand of the House of Lords that research should be politically represented (a part-time minister would have been a better solution). At that level, the greatest danger is that the new advisory committee will be packed with people who echo the government's low opinion of research whose end result is in words, not numbers. The intention that some institutions of higher education should become predominantly teaching institutions is an acceleration of the strategy the University Grants Committee (UGC) has been following for the past two years; what will matter is the degree of flexibility left in the system, the sense in which the new "contracts" for the provision of educational services are applied and, again, the extent to which the management of higher education is put in the hands of people who do not understand what higher education is for, or like. Much the same is true of the plan to set up interdisciplinary research units, on university campuses but separately managed; who will choose the topics, and who will run them? Will ABRC's proviso that the necessary funds should be "new money" be met, or will they be quarried from the budget of the UGC?

Is the new pattern irrevocable? In outline, yes. Like it or not, the present government will hold office for the best part of five years, perhaps for longer. There is no constitutional reason why it should not reorganize civil science as it chooses and no political reason to believe that the parliamentary opposition will prevent it from doing so. For what it is worth, both major parties in the House of Commons share the view that research is for making ordinary people more prosperous. The weakness of the research community and higher education is that they have long since lost most of their political friends. The minister now in charge, the Secretary of State for Education and Science, Mr Kenneth Baker, is probably among the best that they could find, his espousal of the plan to abolish academic tenure notwithstanding. But none of this implies that there is nothing to be done. There is every likelihood that the next session of the British parliament will be overwhelmed by Baker's mammoth education bill. The details, which are malleable, will be crucial.

What can be done (and said)? The first need is that the government should understand that there is no difference of opinion on the issue of whether research should generate social improvement, of which prosperity is a part. The arguments are about how and when. Is a university department whose graduates go on to play productive economic roles always to be written off if



its research does not also generate a sheaf of patentable inventions? Does the British government, and the host of reluctant industrial partners it is forever wishing on the universities, appreciate that industrial development projects tend to cost ten times as much as the basic research from which they spring? Where will that money come from? And may not the implication that the research enterprise has failed British industry be a mirror-image of the longstanding truth that British industry has been uncommonly unenterprising for at least four decades? If the government's animus against the research enterprise blinds it to these questions, especially in its choice of people to run the new arrangements, it will fail, as will the research enterprise itself. Persuading the government of the crucial character of the task it has set itself is the most urgent need.

Otherwise, there are detailed arguments to make about a host of issues which, at ordinary times, would individually be occasions for major public rows. How will contracts for the provision of "higher educational services" be written, and by whom? If tenure is to be abolished, what alternative arrangements will there be to tempt young people into a profession dependent on short-term contracts? Will the government back its apparent acceptance of the view that British secondary education is too specialized for the health of science (and of education generally) with the funds (especially for longer university courses) needed to strike a better balance? What will be done to enable successful institutions, research institutes as well as universities and polytechnics, to flourish? Will the titular autonomy of research institutions, from universities to research councils, be blessed with an understanding that they can be in charge of their own financial affairs? The months ahead will no doubt be noisy with the reverberations of these and other arguments.

The question of morale remains. The greatest damage done to the British research enterprise during the past ten years is that it has been infected with despondency. It would be too much now to complain that people should have been more stoical. How could they have been after a long period of attrition when funds to support good research have been inadequate, when good people have been forced (or tempted) out of their jobs and when opportunities for recruiting able younger colleagues have been too few to give a sense of change? The government's open suspicion that academics in particular and researchers in general are over-tolerant of intellectual layabouts has not helped, nor has its frequently empty rhetoric about the need to foster links with industry. (Is palaeontology the less important now that oil production from the North Sea is at its peak in the spirit in which, three years ago, agricultural research was cut back because of over-production on Europe's farms?) The failure of the natural leadership of the research establishment to protest (in public, at least) at the follies of the past seven years has been more than contributory to this state of affairs.

Nature's position now is simply that the government's decision at last to do something positive is better than a continuation of the long process of attrition that has brought British science to its present state. The recipe for change that has been outlined is full of obvious dangers. Each person's list will be different, but the best strategy for the turbulent months ahead will be to persuade the British government that the way in which it carries through its announced plans must take account of the true nature of higher education and research, which seems only imperfectly to be understood. If the leaders of the research profession can bring themselves to say publicly what they claim they say in the committees to which they belong, that will help enormously to lift the depression at the bench. A large part of the trouble over several years is the willingness of British bigwigs to believe that their influence as members of government committees is undermined if they also say what they think in public, which is the opposite of the truth. But to attempt to persuade the government at this stage to travel to a different goal would merely reinstate the incoherent dialogue of the past few years. □

Professional fraud

The US Congress, making heavy weather of rules to outlaw insider trading, can learn from science.

SINCE the spectacular decision last year of Mr Ivan Boesky, the Wall Street trader, to cooperate with the US Securities and Exchange Commission (SEC) in an investigation of the unfair use of private information to line one's own pockets, financial communities worldwide have been on tenterhooks. Both in London and New York, eminent professional people have been charged with the crime of cheating and found guilty. Some have even been sent to gaol. The sentence in Mr Boesky's case of an agreed plea of guilty to a lesser offence has, however, been postponed. There must be many who ask themselves whether, even now, this accomplished trader is tendering more information for the opportunity to lighten his own burden.

What better time for the US Congress to be asking how the crime of insider trading should be defined? The matter is also made urgent by the hearing by the Supreme Court of the case against a *Wall Street Journal* writer already convicted of improper dealing (advising friends to buy shares whose purchase he was about to recommend); his case before the court will be that he may have broken his newspaper's rules for proper conduct, but that that by itself is not unlawful. (He will probably get off.) Curiously, the Congress which has often investigated fraud in science with much of the zeal it is not devoting to insider trading, has not yet tumbled to the fact that the cases have much in common.

The most evident common attribute is the difficulty of the definition of what amounts to fraud. In science, there are a few obvious reference-points — plagiarism and the fabrication of data — which, if recognized as such, are also recognized to be inconsistent with professional seamliness. It is much the same on Wall Street, or in London, when a partner in a merchant (investment) bank buys up shares in a company on which a professional client of his (or hers) has designs. Everybody knows that cases of that sort are indefensible. The interesting question, though, is where to draw the line between them, the others and honest dealing.

In science, as on Wall Street, there is a natural tendency to cast the net of misdemeanour more widely. Earlier this year, two researchers at the US National Institutes of Health made a convincing case for believing that scientists should not earn credit (in many ways the equivalent of traders' money profits) by appearing only nominally among the list of authors of a scientific publication. Morally, it is wrong, but latterly, the risks have become too high. Better, the moral is, that the person who has done the work should take the responsibility on his own shoulders, but be careful in the course of writing about his supposed achievements to give thanks to those who have helped him professionally.

The principle translates directly into the language in which the US Congress is trying to find a formula for legislation. People who are personally involved in arranging deals between one company and another know they are forbidden to profit from their secret knowledge. So, too, should be those who are professional colleagues in the sense that they give help and advice or provide a working environment. Equally, those who glean information at professional gatherings (in science, meetings) should be forbidden from profiting personally, easily done by declaring an interest at the beginning and implicitly undertaking not to change it. This argues for a definition of insider-trading less extensive in the immediate future than that for which SEC has been canvassing the Congress, but which will ultimately require that even traders should behave professionally. The essence of that condition is that people will have to be able to set personal gain aside for the pursuit of a duty to their clients. Does even Wall Street have the resources to pay for such detachment? Science does not. □

Trials of vaccine against AIDS to begin in humans

- Safety and immunogenicity tests in October
- Other vaccines on the way

Washington

MICROGENESYS has won the race to be the first in the United States to win approval to test a vaccine for AIDS (acquired immune deficiency syndrome) in humans. Last week the Food and Drug Administration gave MicroGeneSys permission to begin clinical trials in October.

The Phase I study will test only the safety and immunogenicity of the proposed vaccine. Eighty-one volunteers — seventy-five homosexual men and six people with no history of risk behaviour — will participate in the trial at the US National Institutes of Health (NIH). Eighteen of the volunteers will be given a substitute injection, to serve as a control group.

Male homosexuals were chosen as subjects for testing the vaccine because AIDS has hit hardest in this section of the population. Homosexuals have often also been exposed to other viral infections at higher rates than in the general population, and so may have different immunological profiles. If the proposed vaccine proves to be safe in the six-month Phase I study, further volunteers will be added in the following year in Phase II, to determine the appropriate dose. But effectiveness will not be evaluated until Phase III.

Because it will not be ethically possible to challenge vaccinated humans with the AIDS virus directly, a long-term prospective study will probably be needed to determine the vaccine's ability to prevent infection.

MicroGeneSys collaborated with researchers at NIH in developing the vaccine, which consists of the gp160 protein from the envelope of the HIV (human immunodeficiency virus), the virus that causes AIDS. Most current AIDS research is concentrated on the envelope proteins because these are thought to be capable of producing the strongest immunological response. MicroGeneSys produces gp160 by recombinant means, using an insect cell-line expression system and a baculovirus vector. Because the vaccine is not derived from whole or live virus, there is no chance that it can cause AIDS.

Anthony Fauci, director of the National Institute for Allergy and Infectious Diseases at NIH, says the gp160 protein evokes high levels of neutralizing antibodies and a good cell-mediated response when administered to small mammals, rhesus monkeys and chimpanzees. But chimpanzees — the only animals other than man capable of being infected with

the AIDS virus — were not challenged by the AIDS virus to determine if the vaccine was effective in preventing infection.

Jorg Eichberg, a leading researcher in evaluating animal models for testing AIDS vaccines, says it is "preposterous to put it in people" before evaluating it fully in chimpanzees. But Fauci replies that infected chimpanzees do not go on to develop AIDS, so that data about the efficacy of the vaccine in chimpanzees may not be transferable to humans.

Another candidate vaccine, which has been developed by the Bristol-Myers subsidiary Oncogen, has been shown to be ineffective in preventing infection in chimpanzees (see *Nature* 328, 721–723; 1987) but is nevertheless expected to be used in trials in humans by the end of the year, according to Oncogen's Shiu-Lok Hu. This vaccine is based on a recombinant vaccinia virus that expresses the genes for the HIV envelope proteins.

Most of the other companies working on an AIDS vaccine are studying gp120, the external portion of the HIV envelope protein. Laurence Lasky of Genentech says the company has tested gp120 expressed in mammalian cells in chimpanzees, but will not discuss the results. Chiron is collaborating with Ciba-Geigy to produce gp120 in yeast, and Repligen has teamed up with Merck & Company to work on gp120 and gp160. Cambridge Bioscience is trying out a variety of approaches, joining Institut Merieux of France to study several viral subunit proteins, a vaccinia virus-based vaccine, and anti-idiotypic antibodies.

Paul Luciw of the University of California at Davis says the bulk of the AIDS vaccine work has been on this protein because it has shown promise in animal studies, and the gp160 protein is difficult to produce. He says there are potential problems with a gp160 or gp120 vaccine. The protein could promote the fusion of T cells in the immune system by binding to the CD4 receptor on those cells. Alternatively, gp160 or gp120 could bind to T cells and cause them to be earmarked for destruction by the immune system, wiping out the very population of cells the vaccine was designed to protect.

Fauci calls these "way-out possibilities", and says there is no indication of immunological aberration in animal tests done so far. Besides, he says, "the only way to tell would be a limited Phase I study in humans".

Carol Ezzell

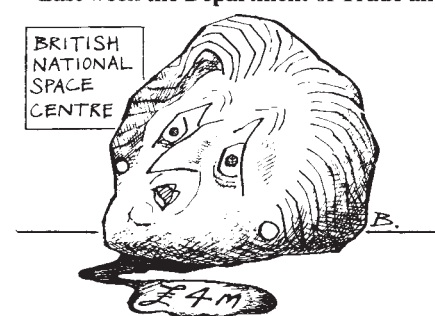
Surprise money for UK space

London

HAVING flatly declared last month that no more money was available for British space research, the government last week produced an extra £4 million, sufficient to keep open Britain's options in the European Space Agency (ESA) research programme.

Roy Gibson, director general of the British National Space Centre (BNSC), the body which coordinates civil space research in Britain, announced his resignation on 4 August when he was told that not only was BNSC's strategic plan, which requested an increase in spending from around £100 million annually to nearer £300 million, being rejected, but that a further £7 million would not be forthcoming to tide over various domestic and collaborative projects until ESA's ministerial meeting in November.

Last week the Department of Trade and



BLOOD OUT OF A STONE

Industry (DTI), BNSC's parent body, announced that an arrangement had been reached with aerospace companies involved in the domestic space programme, notably British Aerospace and Rolls-Royce, whereby the companies would take over the government's portion of the spending, amounting to around £3 million. The main projects affected are studies on the *Hotol* space plane and *Radarsat*, a collaborative venture between Britain and Canada.

The government says that in view of the 'positive' response from industry, an extra £4 million will be provided from DTI's research budget to keep Britain in the running for contracts on *Ariane V* and *Columbus* if the projects are given the go-ahead at ESA's November meeting. Britain's space community sees the extra cash as a small but promising indication that the government is willing to keep its options open until its own review of civil research and development scheduled for October, and the ESA meeting.

The change of mind followed pressure from industrialists and ESA representatives during recent meetings with the government.

Simon Hadlington

British Association adapts to its higher political profile

London

VYING for media attention with the Edinburgh festival (of the arts), this year's gathering of the British Association for the Advancement of Science (BA) has a familiar look. Matters of moment will be aired — with the Right Reverend Hugh Montefiore sure to attract attention by restating the value of chastity in the control of AIDS (acquired immune deficiency syndrome) when the topic is covered in this year's 'BA Debate'. Elsewhere, amateur astronomers and dinosaur watchers are well catered for and local interests are well served — with 'The Spanish Armada and Ireland' a highlight to coincide with the Armada's imminent 400th anniversary.

But the present mood in the scientific profession in the United Kingdom, exemplified by the steady progress of the Save British Science pressure group, has given a sense of purpose to the BA that seemed lacking in the past. Opening proceedings (this year at Queen's University, Belfast) earlier this week, BA president Sir Kenneth Durham set the agenda for a year of campaigning aimed at the "apathy of the general public towards science", which, he says, is "reflected in, and indeed compounded by, attitudes of some businessmen, academics and scholars — people at the front-end of their professions".

The BA's new-found evangelical approach is likely to mean a more 'political' stance from now on. During the recent general election, Durham wrote to the party leaders asking their views on the importance of science to the national economy. The replies were "unenlightening", says Durham, who adds that during the election campaign he heard no party talk about research "except in the most general and vague terms". There were, he says, a few mentions of "the need for more investment in development", but these "missed almost entirely the point that development is dependent on fundamental research".

Industry, says Durham, "has been slow to come to terms with the rate of technical change, and has not played a sufficiently constructive role in linking with education in the past". Durham detects signs of improvement, however, and can even be positive about the much-maligned financial institutions in the City — "the perceived short-term approach of the City and industry to investment is changing and longer-term investment policies are beginning to emerge".

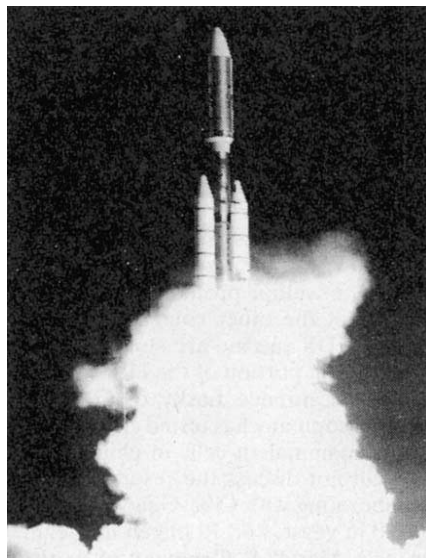
So the public and industry are to be targeted in the BA's campaign — which will depend on additional funding to be sought from industry to support a new

'corporate plan' — but the non-interventionist policies of the government are also to be tackled. Industry should find development ("the most expensive part of the science and technology spectrum") but government, says the BA president, "should concentrate on those research projects likely to be important to the country but unlikely to be funded by industry".

Later this week, during the sessions entitled "Science Audit", more ammunition for the BA campaign should emerge.

Charles Wenz

Voyager's tenth birthday



If other parts of the US space programme are having problems, there is at least one mission NASA (National Aeronautics and Space Administration) can crow about. The space agency is celebrating the tenth anniversary of the two Voyager planetary probes, and although Voyager 1's mission is essentially complete, Voyager 2 still has a planet to visit.

Voyager 1 (above) left Earth on 5 September 1977. Voyager 2 left a little earlier, on 20 August. The two spacecraft visited both Saturn and Jupiter, taking advantage of a rare alignment that permitted the planetary 'tours'. At Saturn, Voyager 1 left the plane of the ecliptic, but Voyager 2 used Saturn to change its velocity and head for Uranus. The Uranus encounter took place in January 1986, and there too Voyager 2 took advantage of local gravity to change its trajectory, this time aiming for Neptune. Voyager's closest approach to Neptune will occur on 24 August 1989.

The Voyager programme was intended to end with the Saturn flybys in 1980 and 1981, but the success of those encounters prompted NASA to keep it going. Through some clever reprogramming of its onboard computers, Voyager 2 has been able to keep up a stream of data, and has already sent back its first picture of Neptune, still some 1,400 million km away. □

Fang Lizhi "no Sakharov"

London

FANG Lizhi, the former vice-rector of China's University of Science and Technology, has repudiated suggestions that he is the "Chinese Sakharov". Last January, Fang was relieved of his university post and expelled from the Communist Party for allegedly having encouraged his students' demands for "bourgeois liberalization". But, as Fang stressed in an interview published last month by the Hong Kong journal *Cheng Ming*, there is an important difference between Sakharov and himself. He, Fang, is not "isolated". On the contrary, since he was expelled from the party he has received many letters "from all directions and from different social strata", letters which, he implied, were supportive.

Fang gave his interview in Rome, where he had been attending a meeting on astrophysics. The fact that, although in disgrace with the party, Fang has been allowed to continue his scientific career (he now has a post at the Chinese Academy of Sciences' Beijing University), is frequently cited by the Chinese establishment as an example of the new democratization of science. (Such claims ignore the fact that Fang has been refused permission to attend conferences in the United States and United Kingdom.)

Fang's own comments on the situation, however, although guarded, are less happy. "If you are repudiated by the party", he said, "you have only the freedom to confess your fault, but no right to refute criticism against you". So far, he said, he has not made any self-criticism, but "you cannot say that there is no pressure on me". Nevertheless, he added, the pressure has been less than "people abroad have expected".

Fang is clearly trying to take an optimistic view of his situation. He believes that by continuing scientific research he can help to hasten democratization. Science, he said, cannot develop without democracy.

China, Fang said, is now facing a "sociological Planck effect". The older generation cannot be expected to change its views, but, as in the case of Planck's constant, the ideas of democracy will ultimately be accepted when the older generation have all died. His own error, he considers, was not that of being a "Chinese Sakharov", but a "Chinese Marco Polo". Marco Polo's descriptions of the Orient, he explained, were disbelieved as travellers' tales. Likewise, he, Fang, had been accused of talking nonsense when he tried to introduce "the ABC of the West" into China. Vera Rich

Rocky road ahead in Senate for NSF budget increase

Washington

THE National Science Foundation (NSF)'s chances of obtaining a large budget increase in 1988 had been looking good, with steady progress in the House of Representatives. But all may now be lost in the Senate. There the appropriations committee is desperately hunting for cuts in total 1988 spending and it seems likely that NSF will be one of the victims.

Eric Bloch, NSF director, is maintaining a bold face in adversity. This week he is mounting a "big push" to persuade senators of the importance of boosting NSF funds and says "I don't want to be optimistic or pessimistic. Nothing is certain until the last vote is in". But it is now quite likely that NSF's budget will remain at the same level as last year.

Bloch remains convinced that support for the long-term goal of doubling NSF's budget remains solid with both the president and many congressmen still expressing enthusiasm for increased spending on science. The first step on the journey, a 16.5 per cent rise in next year's budget, had looked as though it might not be too difficult to make. Despite a few ups and downs in the House of Representatives, Bloch's argument that the nation's economic competitiveness is linked to performance in basic science had proved persuasive and the House Appropriations Committee completed deliberations with a 13.5 per cent increase for NSF.

A few of NSF's priorities were modified. Plans had been laid for thrusts in three main areas: education, including efforts to improve teaching at the pre-college level; improvement in backing for basic research and specialized research facilities; and the establishment of "science and technology centers" where the focus would be on fundamental research that contributes to economic competitiveness. The House Appropriations Committee gave particular backing to pre-college science education, at the expense of research.

Floor action before the full House of Representatives was expected to come soon after the House Appropriations Committee decision (see *Nature* 328, 5; 1987) but was put off until September, at least partly to try to protect the increases from any global cuts that might be arbitrarily imposed in an effort to make the whole 1988 balance. It is exactly that situation that is now being faced in the Senate.

Following a long struggle, the Senate Appropriations Committee last week could find no other way to resolve competing demands for funds from its subcommittees than to ignore its own earlier budget resolution and impose a number of across the board cuts.

The defence subcommittee was protected and the education and the labour, health and human devices subcommittees managed to win their battles for funds. One loser was the HUD-independent agencies subcommittee that is responsible for funding for NSF, the National Aeronautics and Space Administration (NASA) and the Environmental Protection Agency as well as the quite unrelated Department of Housing and Urban Development and the Veterans' Administration. Subcommittee chairman William Proxmire (Democrat, Wisconsin) showed little spirit in the budget battle.

Not much room is now left to manoeuvre in the Senate. Many of the programmes funded by the HUD-independent agencies subcommittee are those that the administration is legally committed to continue. The science programmes of NSF and NASA are far more likely to bear the brunt of the subcommittee's shortfall than are public housing or the Veterans' Administration. But NASA can argue that any cuts could lead to delay or cancellation of the space station — a project to which the United States has already given a degree of international commitment. That leaves bleak prospects for NSF. A decision on how the subcommittee will

New AIDS control recommendations

Washington

To prevent the transmission in health-care settings of the virus causing AIDS (acquired immune deficiency syndrome), the US Centers for Disease Control (CDC) recommend that blood and other body fluids "from all patients" be treated as potentially infective. The recommendations, promulgated by CDC last week and likely to be widely adopted, also urge that in cases where a health-care worker has had parenteral or mucous membrane exposure to such fluids' the source patient be tested for the presence of the AIDS virus. Although the CDC recommendations suggest special precautions to prevent the spread of AIDS virus infection, they also take some pains to demonstrate that virus transmission in health care settings is still a relatively rare event.

Joseph Palca

share out the funds is likely to come at the beginning of September.

The best chance for NSF to pull off a last minute recovery remains when House and Senate get together later in September to work out a joint 1988 budget. House backing for the NSF might well help the NSF to find a higher figure. But whether the final budget will be any larger than this year's will most likely remain doubtful right until the end.

Alun Anderson

New US position threatens to jeopardize CFC agreement

London

THE British government has apparently toughened its stand on the need to protect the Earth's ozone layer by regulating chlorofluorocarbons (CFCs), and is now urging the European Community (EC) to go into international negotiations in Montreal in early September with a position supporting a production cutback of up to 50 per cent.

But there are suggestions that some elements within the US government are trying to scuttle the international ozone agreement by introducing new initiatives into the negotiations, in spite of President Reagan's authorization for a strong US stance on the range of compounds covered and the extent of controls.

One outstanding issue revolves around the percentage of producer countries that would have to ratify the protocol before it would come into force. Some negotiators are concerned that insistence by any party on ratification by countries representing 80 per cent or more of current CFC users could give relatively small users, such as the Soviet Union, power of veto.

Disagreement over the inclusion and

treatment of ozone-depleting halons also poses a potential obstacle. The United States is pushing for the inclusion of specifics on halon control, but the EC says there is not enough information yet for regulations.

Member states of the EC are asking for Single Area Status, or group treatment, for purposes of compliance with CFC reductions, and they argue that there are precedents for this move. But the United States points out that the European countries each vote separately in negotiations on the protocol, and says that insistence on Single Area Status seems to be an inequitable precondition.

Disagreements over these issues threaten to disrupt the progress achieved so far on the Vienna Convention for the Protection of the Ozone Layer, and some participants argue that it is not at all necessary to close all the loopholes as long as the ozone layer is protected.

More than thirty countries are expected to sign a protocol next month, after several years of discussions and negotiations — provided any disagreements can be overcome.

Kathy Johnston

Japanese media scent Nobels in superconductor race

Kyoto

High-temperature superconductivity still has the power to switch on the media (if nothing else). The 18th International Conference on Low Temperature Physics opened in Kyoto last week under the full glare of television cameras and lights. And on the second day a Japanese researcher set the cameras rolling again with a claim of superconductivity at 338 K (65 °C) in a new type of copper oxide.

The conference, sponsored by the International Union of Pure and Applied Physics, the Science Council of Japan, the Physical Society of Japan and the Japan Society of Applied Physics, has been swamped by contributions on high- T_c superconductors, with over 200 of the 1,000-odd presentations (talks and posters) in the field. And nearly half (41 per cent) of the ¥120 million (\$800,000 million) cost of the six-day conference has been put up by Japanese industries interested in the development of the new superconducting materials.

At the opening ceremony on 20 August, 20 recipients of the Fritz London Memorial Award found themselves at the centre of attention befitting Nobel Prize winners. John Clarke from the University of California won recognition for his pioneering development of SQUIDS (superconducting quantum interference devices) and SLUGS (superconducting low-inductance galvanometers), work he began at the Cavendish Laboratory of the University of Cambridge in the 1960s. And Jun Kondo of Japan's Electrotechnical Laboratory was given an award for his explanation of the resistance minimum in metallic systems with magnetic impurities. But the television cameras and floodlights were for J.G. Bednorz and K.A. Müller, who shared the third award for their discovery, last year, of superconductivity in Ba-La-Cu oxides.

The television crews got another shot at the stars of high- T_c superconductivity on the following day when five of the world's leading researchers presented review papers. Although nothing particularly startling was said, the photographers rushed forward *en masse* at the start of each talk for their allotted one minute of picture-taking.

V.L. Ginzburg of the Academy of Sciences of the USSR, who was given an unscheduled slot at the end of the review session, waved the press aside, pointing out that he could not possibly say much of significance in three minutes. He then chided his colleagues for being swept away in the "mania" of high-temperature superconductivity (although he noted it had provided him with his first opportunity

in 44 years of research to attend the low-temperature physics conference) and he appealed to young researchers to avoid "unhealthy conflict".

But at the afternoon poster session on high- T_c superconductors, the mania continued, with packed aisles more reminiscent of a Tokyo station during rush hour than a scientific conference. And the television crews continued to trail the review speakers, asking questions like "Do you think you will get a Nobel Prize?"

Dr H. Ihara of the Ministry of International Trade and Industry's Electrotechnical Laboratory grabbed the most limelight, however, when he announced during Friday evening's short-talk session that he had observed zero resistance in the new Y-Sr-Ba-Cu oxide ($\text{YSrBaCu}_3\text{O}_{8-x}$),

$\delta > 0.2$) at 338 K. Flash bulbs lit up the hall when Ihara presented a resistance-temperature curve that plunged to zero ($10^{-8} \Omega\text{cm}^{-2}$) above room temperature and again when he showed a curve of magnetic susceptibility with a slight break in slope at almost the same temperature (340 K).

But audience reaction was sceptical. Zero resistance, although stable for ten days in one of the two samples (one day in the other), was observed in only about 0.01 per cent of the 9-mm-diameter sample in a zone confined to the perimeter of the disk. Some suspect that the isolated observations of zero resistance could be an artefact of minute cracks in the sample that open and close due to thermal expansion and contraction.

Yet, despite the scepticism, the telephone lines in Kyoto were set abuzz that night, not only by the press but also by researchers calling home to their laboratories with the recipe for the latest oxide.

David Swinbanks

Japan's education reformers rallying round the flag

Tokyo

JAPAN'S National Council on Education Reform has ended its three years of deliberations without finding a clear solution to the education system's biggest problem of intense competition for places in elite universities. It does propose some university reforms, principally that the universities should change the start of their academic year from April to September to fall in line with other countries, and that the research system should be altered. But no date is given for the former change and the latter is left largely to a yet unformed university council.

The council's final report criticizes the education system for its failure to encourage individuality and bring forth scholars of international standing. But the deep division over how to tackle these problems reflects the conflicting views of traditionalists, aligned with the Ministry of Education, Culture and Science, and those seeking radical reform. The report stresses, on the one hand, the need to foster children with liberal and autonomous minds who can develop their creativity and, on the other, the inculcation of patriotism and respect for social norms.

Overall it would seem that the traditionalists have gained the upper hand. Since the council brought out earlier versions of these proposals, the Ministry of Education, Culture and Science has moved to encourage schools to display the national flag and enforce singing of the national anthem. Some school districts have fallen into line. Elsewhere there have been protests, especially from those who remember pre-war nationalist days.

In Okinawa, an irate mother ripped down the flag at her child's school entrance ceremony to a mixture of applause and boos from the audience.

The intense competition for entry to top universities which forces children to spend long hours at evening cramming schools seemed beyond the council's ability to tackle. Change is likely to come only if government ministries and large companies stop giving preference to graduates of elite universities. All the council could do was to call for local education boards to "look into" cram schools and "to try to solve problems, if there are any". Support is given to the idea of life-long education and the council chairman, Professor Michio Okamoto, hopes that improved educational opportunities for working adults might do something to ease the competition for university entrance. So far, however, university faculties that run public lecture courses have often run into opposition from colleagues who are competing for scarce resources.

In the universities, a call is made for shorter-term appointments to replace the lifetime employment system. Encouragement is also given to a review of the grant system, the increase of funds for research, and an improvement of the postdoctoral fellowship system. At present, only a few hundred postdoctoral fellowships are available. Despite over 2,000 hours of debate by the council, many of the universities' key problems, such as the chronic lack of technicians, were left unresolved. A new "university council" is planned to take over where the reform council left off.

David Swinbanks

Food irradiation legislation due despite public suspicion

London

THE British government is expected soon to legalize the irradiation of food as a means of sterilization, bringing Britain into line with 29 other countries, including most of its partners in the European Economic Community (EEC). Treatment of food with ionizing radiation is at present prohibited in Britain, as it is in West Germany. All the other EEC states allow the process for certain specified foods.

Last year, the independent Advisory Committee on Irradiated and Novel Foods (ACINF) presented the British health minister with a report on safety aspects of food irradiation. Its overall conclusion was that there was no evidence to justify the continuing ban on food irradiation, provided that specified dose limits were not exceeded.

Shortly after parliament resumes in October, the government is likely to



announce legislation to permit the irradiation of certain foods, the three most likely candidates being spices, grain and poultry. The European Commission is also considering the subject, in the context of its 'internal market' policy, which is intended to make possible the free movement of foodstuffs between member countries by 1992.

The Commission is preparing a draft directive on irradiated foods. If, as seems probable, the directive is adopted and if Britain and West Germany have not by then legislated, they would be obliged to follow the directive unless, within 18 months, they could provide evidence of toxicity arising from irradiation.

The ACINF report recommends an overall average dose limit of 10 kilograys, sufficient to delay ripening or kill pathogenic microorganisms without rendering the food radioactive.

Opponents of the method point to potential difficulties in the enforcement of controls. There is at present no reliable means of identifying the small quantities of novel radiolytic products formed within

irradiated food, making it almost impossible to determine efficiently if a food has or has not been irradiated, or how many times. In the absence of such tests, controls would need to be enforced by strict documentation at irradiation plants, a method, say the opponents, fraught with the potential for abuse.

Other commonly expressed fears are that the process could cause the selective survival of pathogens, formation of bacterial or fungi toxins, or the creation of resistant or harmful mutants. Despite ACINF's assurances that 30 years' investigation has produced no evidence of such effects, in the wake of Chernobyl a suspicious British public is bound to ensure that legislation will be given a rough passage.

Simon Hadlington

Extra mileage in ethanol for Chirac

Paris

FRANCE continues to support plans to convert part of its grain and beet mountain to ethanol, a petrol substitute, despite rejection of the idea by its partners in the European Economic Community (EEC). Although all the evidence shows that the scheme would be uneconomic, Prime Minister Jacques Chirac is anxious not to offend powerful agricultural lobby groups before the 1988 presidential elections.

At present the export of surplus grain to the developing countries is subsidized by the EEC. Farmers have argued that this subsidy could directly be converted to an ethanol production subsidy.

But a study commissioned by the European parliament shows that ethanol costs four times as much to produce as petrols, or other fuel additives. A much larger subsidy than that required to export the surplus grain would be needed to turn the grain into alcohol. Without help from its European partners, such a scheme would cost the French government FF3,000 million a year.

But there may even be some political mileage to be gained from going against the rest of the EEC and supporting the scheme in France. A 1983 law allows French petrol to contain up to 5 per cent ethanol. Over the past two months, Chirac has reduced the tax on ethanol and encouraged its use. That pleases the agricultural lobby and, by a remarkable sleight of hand, it costs the government nothing.

Because of its low calorific value, adding ethanol reduces fuel efficiency and as much comes back from value added tax on the extra petrol sold as is given out in concessions.

Peter Coles

Britons beating CERN drum

London

BRITISH employees at the European high-energy physics laboratory (CERN) at Geneva are taking early action to influence the British government's decision, expected later in the year, on its continued membership of the 12-member organization.

In a statement put out earlier this week, Eifionydd Jones, chairman of what he called the CERN-UK Working Group, says that British scientists working at CERN, of whom there are 334, believe that the British government's decision on whether or not to continue as a member is imminent, and that "there is a very real threat" that Britain may withdraw from membership.

A large part of the accompanying case for continued membership is the argument that developments in high-energy physics have stimulated the development of accelerators for treating cancer patients (of which there are now more than 1,000 world-wide), similar facilities for the production of short-lived radioactive isotopes and machines for carrying out ion-implantation of the kind used in the manufacture of microcircuits.

The CERN-UK working party claims the turnover of Oxford Instruments Ltd. (quoted as £55 million a year), a successful manufacturer of superconducting magnets, as one of the benefits of high-energy physics to the outside world. These examples are cited as proof that basic research and technology develop hand-in-hand.

The working-group's "FACT PACK" also includes what is in effect an appeal to British manufacturers to take competitions for CERN contracts more seriously. While CERN goes to considerable lengths to arrange that competition for tenders will be equitable, basing its awards on factory-gate and not delivered prices, British companies have not usually recovered what they might consider to be their fair share of CERN's capital spending.

An important part of the working party's recommendations is that those who receive its information-pack should "speak out to all those with influence and a voice in public affairs". The group's statement says that the idea of the United Kingdom's withdrawal from CERN "seems incomprehensible to our European colleagues".

How correct may be the working party's estimate of when the British government will decide about continued membership is not clear. The most likely timing of that event is just before the next council meeting of CERN, fixed for December. □

Winds of change blowing hard in British science museums

London

DECLINING support from the government is forcing Britain's science museums to think hard about ways of reducing costs and attracting funds from alternative sources.

The government is loosening its hold over the museums in an effort to encourage them to forge closer partnerships with the private sector. The British Museum (Natural History), BM(NH), was until a month ago under the aegis of the Department of Education and Science, receiving its funds (£11.44 million last year) from the science budget. On 1

Last year, the Science Museum received £8.8 million from the government, with earnings from publishing, retailing and hiring of rooms totalling £879,000. Sponsorship from industry netted almost £1 million. Cossons believes there is much scope for increasing earnings from non-government sources. But many of the museum's curators are suspicious. Cossons' pledge that he has "no intention" of introducing compulsory redundancies is not regarded as a sufficiently strong guarantee of job security. Several members of staff have been refusing to cooperate with external management consultants brought in to find the best way of implementing the changes.

So far, Cossons has refrained from recommending the introduction of admission fees to the museum's public galleries. He is, however, keeping a close eye on BM(NH) next door, which started charging visitors on 1 April this year, and whose attendance trends traditionally closely reflect those of the Science Museum.

BM(NH) says that it is too soon to gauge the effect of the charges, although admissions do not seem to have changed significantly so far. BM(NH) is pinning its hopes on the success of the new charging policy which is calculated to generate £1.6 million annually by 1990-91. The first formal appraisal of the effect of admission fees is due towards the end of this year. The museum could find difficulty in convincing opponents of the new policy that an estimated 20 per cent drop in attendance figures was expected simply by virtue of the inaccuracy of assessing the numbers of visitors before charging was introduced.

Because curation and research swallow by far the largest proportion of BM(NH)'s resources (41 per cent in 1985-86), the museum's management is trying to lessen the financial burden they impose. If further manpower cuts are necessary, the scientific staff will bear the brunt. The management is also considering ways of marketing the abundance of scientific expertise.

Like the Science Museum, BM(NH) is embarking on a fierce marketing programme. With industrial sponsorship unlikely to be a big money-spinner, most efforts are being concentrated on giving the museum a higher public profile, aimed at pulling more paying customers through the turnstiles. Renting rooms for parties, something that has attracted the attention of the national media, represents a further, small source of income.

When BM(NH)'s new director, Neil Chalmers, replaces Ronald Hedley next year, he will be taking charge of a much-changed institution. **Simon Hadlington**

Academic networks

Taking stock

Munich

STILL in its infancy, the European Academic and Research Network (EARN), a computer communications network, is now embarking on enforced independence without the comforting support of computer giant IBM. EARN must learn to live with the potentially disastrous tariff policies of some European telecommunications monopolies, earning at least enough income to replace the \$15 million provided by IBM over the past four years. But director Dennis Jennings is confident that EARN will stay the course.

Since 1984, EARN has connected more than 2,000 computers in 20 countries, largely with the help of IBM's determination to overcome wildly different standards and attitudes in the host countries. The network now includes Israel, the Ivory Coast and Iceland. All members are also connected to US research institutions through BITNET, a US computer network.

EARN's success is measured by its growth. In West Germany, for example, EARN traffic has doubled every 10 months since the beginning. The West German data lines, which carried 2,400 bits per second (b.p.s.) at the outset, have been largely replaced by lines carrying 9,600 b.p.s. and may soon be upgraded to 64,000 b.p.s., already available in many other European countries.

But who will pay the costs when IBM support runs out at the end of 1987? EARN directors decided at Nice in May to adopt the "BITNET model" to fund the lines connecting the various countries. Under that scheme, each user pays for the line leading to its neighbour; the members share the cost of one line to the United States and the other line is paid for by West Germany.

Because of huge discrepancies in the cost of the international lines, however, several countries are considering shifting their lines out of high-priced areas such as West Germany (see opposite) and Switzerland. The United Kingdom, for example, is contemplating a shift of its lines from Geneva to Montpellier, France. West Germany may lose all but one of its five international data lines because its charges are so high. These shifts are not expected to hurt EARN's performance.

Nor do most countries expect difficulty with the transition to self-management. Networks such as SWITCH in Switzerland, JANET in Britain, SURF in Holland and REUNIR in France will continue to belong to EARN. The only problem will be convincing funding bodies to give support to the network. But EARN director Dennis Jennings is confident that all countries have established a

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Suresh Karanth/The Independent

A wedding reception among the dinosaurs now possible at BM(NH).

August, BM(NH) was brought into line with the other main museums and passed over to the Office of Arts and Libraries.

On 1 April, funding arrangements for BM(NH) and its neighbour, the Science Museum, were changed, with the government money coming as grant-in-aid, releasing the museums from many of the Treasury's annuity restraints.

Furthermore, the government is shortly to hand over the responsibility for the museums' buildings and estates from the government's Property Services Agency to the museums' trustees. The sums are not inconsiderable — in 1985-86 PSA was given £4.8 million to spend on BM(NH).

At the Science Museum, the new director, Neil Cossons, appointed 18 months ago, is making his presence felt. He hopes to bring the four curatorial departments under the single roof of a 'collections management' department. Research priorities will be redefined by a new head of research, with responsibility for the library service, and a separate marketing division is to be created.

... academic networks

... without IBM support

source of funds, at least for 1988.

One obvious area for improvement is in EARN's carrying capacity. European researchers are envious of the new US network NSFnet operating at 1.544 megabits a second. In Europe, Jennings says there are "no plans" to upgrade EARN beyond 64,000 b.p.s.

One of EARN's most important tasks in the transition is to tell whether demand justifies such massive increases in capacity. In some fields, such as computer-aided design, the need already exists; but in areas like the humanities demand will depend on how well EARN can sell itself.

Steven Dickman

Munich: Expensive survival

THE prospect of IBM's withdrawal as a supporter of BITNET has dramatically drawn attention to the high cost of long distance communications in West Germany. Because of a new volume-based tariff structure for international leased data lines, the European network EARN plans to shift as many lines as possible out of West Germany. But two of eleven *Länder* plan to underwrite the cost of data transmission for research.

The dispute over tariff policy concerns five EARN lines terminating in West Germany. Since the partners sharing the lines also share the cost, the effective 100 to 500 per cent increase in price makes them unattractive. EARN director Dennis Jennings calls the tariff policy "quite outrageous". West German EARN director Michael Hebggen, of the University of Heidelberg, agrees. He says all the lines may have to be cut if the ministry does not change its policy. Others believe that, while the new policy may cause problems, academic network links will not be cut.

The postal ministry defends the high tariffs by citing the high quality of the West German telecommunications network. Based on the "equal treatment principle", it also says that scientists should be no more entitled to lower rates than anyone else. But the postal ministry's argument that researchers should look elsewhere if they want cheaper service seems cynical in the light of its effective monopoly.

Eventually, federal and state governments may step in to help. There is a two-year-old computer investment programme that may be a model for joint financing, according to Alfred Kupllmer from the Deutsche Forschungsgemeinschaft (DFG). The immediate problem is to take over from IBM the costs of the lines leading to the 24 computers (out of a total of 170) forming the "backbone" of the West German network. With federal and state help, West German researchers are

unlikely to suffer much from IBM's withdrawal.

Steven Dickman

Washington: Networks multiply

COMPARED to Europe, US computer networks have it easy. Competition in the telecommunications industry in the aftermath of the break-up of AT&T (American Telephone and Telegraph) has produced an overcapacity of high-speed data lines, and therefore nominal transmission costs. US networks have also become more used to being self-supporting.

That being said, all is not plain sailing for networks in the United States. A national debate is under way over who should pay for a 'backbone' for NSFnet, a national research network just starting to operate. At present, NSFnet is moving about 60 million packets a month through its six-node backbone to its collection of regional networks, campus networks and supercomputer centres. Some argue that NSF should include a line item in its budget to pay for the system, but others would prefer that it be supported as part of the indirect cost of research grants. Another possibility is for the regional networks themselves to pay for the backbone through revenue from their members.

The popularity of networks in the United States has created its own set of problems. Different protocols and communications standards has created the potential for a Tower of Babel, making network interconnectivity a nightmare.

A development expected shortly should help to rationalize some of the problems in sideways communications among networks. BITNET will be merging with CSNET. Both have been popular with the research community, and both are self-supporting from user contributions. CSNET began as a network for computer sciences, and BITNET began life as a general academic network.

Congress has requested that the federal government come up with a plan for the future of scientific computing in the United States. That report has now been essentially completed, is under review within the White House, and should be delivered to Congress later this year.

Joseph Palca

London: EARN poor relation

NEGOTIATIONS are proceeding in Britain over the fine print in a formula for a new funding arrangement for the EARN computer network. Britain has agreed to share the costs of the system with other European users, to the tune of £35,000, which would cover the cost of the leased telephone line to CERN in Geneva as well as a proportion of the satellite link between Montpellier and New York. British network users will also

need to find another £35,000 to cover staffing and overheads.

A formula has been hammered out calling for major funding bodies, such as research councils, to pay a proportion of the costs based on their use of the network over the last six months. Each user group within the research councils would be assessed under an arrangement that EARN's director in Britain, Paul Bryant, admits is "a bit rough" since it is based on machine time rather than individual users.

Provided the research bodies agree to pay a proportion of the computer network's costs, the new funding arrangement should make no difference to British scientists who use the network. If any of the groups refuse to join in, their members will not be able to take part in the network.

The funding formula has won the approval of the Computer Board for Universities and Research Councils, which finances computers in universities. Individual universities within the EARN network will meet to discuss the new formula in November, but it is considered they have no option but to agree. A new funding strategy is under consideration for 1989, when the European network will change to a more modern design.

EARN has not taken off in Britain the way it has in some other countries, partly due to an initial fear that it would undermine the local Joint Academic Network, JANET, and partly because EARN uses older and therefore less attractive technology.

Kathy Johnston

Paris: REUNIR ready

FRANCE's academic computer network will still function next year when IBM ceases paying for its leased telephone lines through EARN. One of the first countries to be involved with EARN, France has recently become a major link in the network with the choice of CNUSC, the national computing centre at the University of Montpellier, as the site for the European high-speed satellite link to BITNET at the City University of New York.

With the withdrawal of IBM funding for EARN, France's major national academic network, REUNIR, will take over a large share of the financial responsibility. Costs will be split jointly between individual sites (universities and research centres), who will pay a subscription towards running costs, and REUNIR itself, which will pay for global fixed costs, such as renting the data lines. REUNIR's budget comes out of the annual computing budgets of its founder-members, the state-funded research organizations.

A spokesman for REUNIR explained that IBM's funding of EARN played a significant pump-priming role and has enabled the networking infrastructure to develop in France, showing that there was a "considerable need" for such a service.

Peter Coles

Plight of UK science

SIR—The leading article on the report of the Advisory Board for the Research Councils (ABRC) on university research (*Nature* 328, 279; 1987) was, to say the least, inadequate. With friends like this, who needs enemies?

The dull acquiescence of the ABRC to the British government's jaundiced view of the universities in general and basic research in particular is bad enough without *Nature* weighing in with a bland, powder-puff commentary. Indeed, it is difficult to remember any leading article in *Nature* over the past few years that has bared any real teeth to the Department of Education and Science. No doubt *Nature* will soon be telling us again in veiled terms that the abolition of tenure is no bad thing.

So what does the ABRC propose? Answer, exactly what the government wants to hear. A small elite of 'core staff' (no doubt enjoying the emasculated interpretation of tenure that Mr Kenneth Baker will soon be thinking up) in a small number of research centres that will entertain 'visiting teams' of workers 'seconded' from their home institutions. This exercise in reality amounts to no more than moving around the remaining bits of furniture at maximal inconvenience for a minimum of benefit. If the ABRC, and *Nature* for that matter, thinks that this dog's dinner of ideas will bring expatriates rushing back from the United States and elsewhere and encourage postdocs to linger a little while longer in basic research, they are whistling in the wind. It is time for those with influence and authority in our scientific community to stop doing the government's dirty work and to say loud and clear 'enough is enough'.

The truth is that science has surrendered to mammon. The government has no patience with anything that cannot be exploited for profit and basic research is a runt in the government's latest litter of 'radical' proposals. As one report follows another in depressingly similar style, with the usual hackneyed clichés 'economic realities' 'need to rationalize', 'become cost-effective' and so on, one is reminded of the scorn the poor bloody infantry in the First World War had for the generals safely back at headquarters. With the notable exception of Denis Noble's Save British Science initiative, our scientific generals have been deafening in their silence. Come on you generals, give your troops the lead they so desperately need otherwise one can only suspect that you are keeping your heads down for a quiet life, and why not? For many, their research careers are over and jolly nice ones they had no doubt, in the decades before this one.

If the government will not listen, those

in the scientific establishment must be prepared to rise to the challenge. If they are too tired to fight, then they should be honest enough to make way for others. It would be nice to think that the challenge will be accepted, even at the eleventh hour.

D. A. W. GRANT

*St George's Hospital Medical School,
Cranmer Terrace,
London SW17 0RE,
UK*

SIR—I have been shocked in recent months by the attitude of leading articles in *Nature* to the government's plans for science in Britain. It is astonishing how rapidly the focus of this argument has moved in the direction the government wishes. An article containing such proposals as the leading article "World's end or its beginning?" (*Nature* 328, 279; 1987) would have been amazing only a year ago and provoked a storm of protest whereas now it apparently provokes only restrained agreement.

I believe the whole debate stems from a false premise, which is that British science is a failure. This seems to me to be so far from the truth as to be almost ridiculous. British science is probably second only to that of the United States in terms of both quantity and quality of output (*Trends Neurosci.* 10, 105; 1987), and this despite a much lower proportional funding. That sounds like a high degree of 'efficiency' to me, especially considering the starvation of funds and low morale of recent years.

It is unfortunate that British science conceded the argument with the government at the very beginning by allowing the accusations of 'inefficiency' and 'failure' to go by default. Once these points were allowed, the present radically destructive policies followed with horrible inevitability.

BRUCE CHARLTON

*MRC Neuroendocrinology Unit,
Newcastle General Hospital,
Westgate Road,
Newcastle upon Tyne NE4 6BE, UK*

Easier said than done

SIR—The widespread use of the phrases *in vivo* and *in vitro* in the scientific literature is undoubtedly due to the fact that much is said with few words. Moreover, the situations referred to in the two cases are encountered frequently. We propose a new and similar expression that corresponds to a frequent occurrence in the planning of experiments.

The need for such an expression was brought home recently when one of our students commented that her senior thesis had become much more complicated and

difficult — when she started carrying out experiments — than it had seemed it should be, based on the planning stages alone; on paper, the idea had appeared simple; in execution, it was much more difficult.

In the belief that this widespread situation can be summed up succinctly, we suggest the phrase describing thought experiments might be *in charta*, signifying 'on paper' and further, that the contrast might be forcefully expressed in *charta facile, in vitro difficile* — that planning something is often easier than carrying it out. The phrase sums up much of the frustrations of research.

C. TYLER BURT

*Laboratory of Molecular Biophysics,
National Institute of
Environmental Health Sciences,
PO Box 12233,
Research Triangle Park,
North Carolina 27709, USA*

JOHN HANCOCK
FREDERIC PEACHY

*Reed College,
Portland,
Oregon 97202, USA*

King Faisal prizes

SIR—The article about the 1987 King Faisal International Prizes (*Nature* 326, 118; 1987) leads with the curious allegation that the history of the prizes has been "somewhat shaky". The independent scholars who control the jury process have occasionally decided that none of the submitted nominations in the designated topic for a particular year really meets the foundation's stringent qualifications. Rather than perceiving that as "shaky", one might take the view that the foundation and the scientists involved in the prize selection process are attempting to maintain the quality and selectivity that the foundation has attempted to build over the past decade.

At the end of the article an inference is made that our selection process has been prejudiced on the matter of sex, religion and race. We have never taken into account any of these factors because we do not control the jurors, whose members are from every part of the world. Dr Michael Field was the co-winner of the King Faisal International Prize in Medicine in 1984 and, as he has pointed out (*Nature* 327, 96; 1987), he is Jewish.

We too regret the absence of female winners. As the international jurors consider only the submitted nominations, however, the blame might better be attributed to the many international institutions that actually do the nominating.

BANDAR BIN SAUD BIN KHALID
*King Faisal Foundation,
PO Box 352,
Riyadh 11411,
Saudi Arabia*

Correspondence not yet for closing?

Special relativity is the most familiar graveyard of general understanding. But can general understanding be advanced when discussion of particular points is suspended?

WHEN can a controversial topic fairly be held to have run its course? This question arises in the management of any correspondence column even when the contributors are as uniformly discerning, and as free from the temptation to grind parochial axes, as those who are good enough occasionally to write to this journal. Sometimes, it seems that correspondents have said everything it is possible to say about a topic, that the chance that one or other of two disputant sides will change its mind has shrunk to zero or that people are about to move from intellectual argument to personal abuse. Then, editors feel free to wash their hands of the business with the somewhat pompous declaration "This correspondence is now closed".

Just that has just happened in *Physics Bulletin*, the monthly news journal of the British Institute of Physics, which in February stumbled into what promised to be an entertaining and even instructive vein of correspondence by publishing a brief article "Energy has mass" by Sir Hermann Bondi and C.B. Spurgin (*Phys. Bull.* 38, 62; 1987). The declared objective was to rid special relativity of a "common misunderstanding". The next step, in April, was a letter of dissent from Sir Rudolph Peierls (*ibid.* p.127) which must have excited readers to look forward to fireworks ahead.

In the event, the outcome is anticlimax; Norman Dombey, Sir William McCrea and John Rousseau say at some length that Bondi and Spurgin have not dispelled confusion, but have added to it. Bondi and Spurgin confess that that may indeed have been the case and the editor of *Physics Bulletin* says that the correspondence "has now been closed". The past tense suggests a greater degree of finality than usual, even a greater degree of finality than readers of the journal will welcome.

What can be the hare run so quickly to ground? Citing half a dozen textbooks and a schools examination question to illustrate what they describe as "the serious nature of this misunderstanding", Bondi and Spurgin originally set out to show that most people misinterpret Einstein's equation $E = mc^2$ as a proof that "mass and energy are interconvertible". One of the worst offenders is the Smyth Report, the document that in 1946 provided the first rattling account of the way in which nuclear fission functions, but which rested much of its case on the

assumption that mass can be sometimes turned into energy. This, Bondi and Spurgin said, is not the case. On the contrary, energy and mass are governed by separate conservation laws which ensure that the amounts of energy and of mass in a physical system are strictly constant. Specifically, "students should be taught that:

- (1) energy has mass;
- (2) energy is always conserved;
- (3) mass is always conserved".

What Einstein's equation shows, Bondi and Spurgin say, is merely that energy has mass, or inertia, a point they illustrate by calculating the difference between the mass of a moving particle and its rest-mass, which is the relativistic kinetic energy divided by c^2 , or $\frac{1}{2}mv^2$ when v , the velocity, is small compared with c , the velocity of light. Another way of putting this is to observe that a photon, which has zero rest-mass, plainly has inertia because, when stopped, it can transfer momentum to an atom.

So how did misunderstanding come to pervert discussion of Einstein's equation? Bondi and Spurgin said that "simplified popular accounts of nuclear fission processes" may have been responsible. True, they say, the combined mass of a ^{235}U nucleus and a neutron (needed to stimulate fission) is greater than that of the products, while "a very considerable amount of energy seems to have appeared from nowhere". But those who have been putting about the heresy that this is an illustration of the conversion of mass into energy have either forgotten or malevolently concealed the circumstance that energy has mass E/c^2 which, when added to the combined mass of the fission products, saves the appearance and the integrity of the principle of the conservation of mass.

Peierls' comment is the more direct, amounting merely to the assertion that it is only a matter of convention, or of convenience, whether the kinetic energy of the fission fragments (which becomes the heat produced by nuclear reactors) is counted in Einstein's equation as mass or as energy.

The difficult case is that of the annihilation of a positron and an electron, most simply supposed at rest. The end-product consists only of photons of zero rest-mass. Would it be more appropriate to say that the final state has energy, or that it has mass (which is equal to the measurable

energy of the photons divided by c^2)?

Dombey, McCrae and Rousseau go further, pointing out that the tangent vector to the world-line of a particle by which its evolutionary history is described, often called the momentum four-vector or the four-momentum, has a fixed length equal to the rest mass (multiplied by c), that the time-like component of the four-momentum (for a particle, the mechanical energy) depends on the reference frame and on its velocity but that the mechanical energy of a system of immutable particles is conserved within a single frame of reference, and so on. Concerning the question of the annihilation of positrons and electrons, they ask directly "why Bondi and Spurgin protest at this being described as the conversion of mass into energy".

To be fair to everybody, the answer they get is not particularly satisfactory. Bondi and Spurgin, while acknowledging that they "may have started more of a discussion, with more of a danger of confusion, than we had realised", go on to say that the differences between them and their critics "are differences in terminology and the perceived needs of readers". They say that they still believe their quotations from other people's textbooks "were only too liable to create a wrong understanding" and that their separate listing of the conservation laws of mass and energy "no more meant they were separate" than "stating such a law in English and in French means that there are two separate laws".

So, in the end, it seems that everybody is at one. "If we wrote the article again", it would have been emphasized that "mass" means nothing but "measure of inertia". No doubt, the authors would also have taken greater care to rid their text of emotive headlines such as "exploding myths" and statements that students should be "urged to avoid such terminology as 'the equivalence of mass and energy'". In any case, the upshot is that there is only one conservation law, although it is far from clear whether people "to whom the distinction between rest mass and inertial mass is less obvious than it is to, say, particle theorists" can be trusted with that knowledge. "We can only ... hope that ... the clarity will result that we are all so keen to create."

With such a goal, how could such a correspondence be closed or, worse, "have been closed".

John Maddox

Superconducting ceramics

Does magnetism hold the key?

V.J. Emery

THE fundamental theoretical problem posed by the new high-temperature (high- T_c) superconductors is to understand the microscopic nature of the superconducting state and the source of the electronic interaction that gives rise to it. BCS theory¹, the universally accepted explanation of superconductivity in metals, makes two logically independent assumptions: that below the critical temperature T_c there is a collective state of pairs of electrons, and that the interaction responsible for it is produced by polarization of the lattice. Both have been called into question by the high transition temperatures and unusual physical properties of the oxide superconductors.

Pairing itself is not in doubt: tunnelling and flux-quantization experiments^{2,3} have clearly demonstrated the existence of entities with charge $2e$ (twice the electron charge). It is possible that they are 'real-space' pairs that are present at relatively high temperatures and that T_c signals a Bose condensation to a state which is superconducting for much the same reasons that liquid ⁴He is a superfluid. On the other hand, the BCS view is that everything, the pairing and the formation of the collective state, takes place at T_c . Most theories could accommodate either possibility. Indeed, the distinction is less pronounced for the oxide superconductors because the relevant electrons move mainly in CuO_2 planes so that, unlike a fully three-dimensional problem, it is not necessary to have a particularly strong attractive interaction to produce real-space pairs. It is never necessary to have a strong attraction for BCS theory to be applied, so both approaches can provide a useful starting point for the oxides.

In general, real-space pairs are more likely to dominate at low carrier density, but localization of the electrons may prevent superconductivity altogether. Real-space pairing would be indicated if, for example, the energy to bind a pair is much larger than predicted by BCS theory. This could be shown by infrared-absorption or tunnelling measurements. On balance, experiments to date appear to favour the BCS picture, but not unequivocally, and advocates of either point of view can find comfort in the literature.

Exchange processes

Whatever the nature of the pairs, they cannot exist at all without an attraction to overcome the coulombic repulsion between electrons. Usually it is assumed that this comes about when an electron polarizes its surroundings to produce an

environment that is energetically attractive to another: in the language of quantum mechanics, the electrons exchange an excitation. The polarized medium (and the corresponding excitations) might be the lattice (phonons), external charges (excitons), the spin density of the superconductive electrons (ferromagnetic or antiferromagnetic spin fluctuations, magnons) or the charge density of the superconductive electrons (plasmons or charge-density waves). All these have been suggested as mechanisms of high- T_c superconductivity. An important feature of these exchange processes is the possibility of retardation. Typically, the interaction to which they give rise is weaker than the coulombic repulsion, but it is effective because the polarized medium is slow to relax, and the second electron feels its influence after the first one has moved away. By the Heisenberg uncertainty principle, timescales are inversely proportional to the typical energies — for example, the Debye energy of phonons or the Fermi energy of the electrons. Curiously, it seems possible that the high- T_c superconductors may not take advantage of retardation. According to Hall-effect measurements⁴⁻⁶, it seems that T_c is roughly proportional to the carrier density n_c or, equivalently, to the Fermi energy. This result is consistent with BCS theory, provided that the Fermi energy is sufficiently small, in which case retardation is not effective. It could also be explained by real-space pairing, for which the Bose condensation temperature is approximately proportional to n_c .

Phonon exchange is the mechanism responsible for superconductivity in all metals, with the possible exception of organic and heavy-fermion systems. It has been argued that phonons could never give T_c s as high as those observed routinely in the oxides. Weber⁷, however, finds that the electrons couple strongly to oxygen breathing-mode vibrations in which the four oxygen atoms surrounding a copper atom move alternately inwards and outwards. Because of the strength of this coupling and the low mass of oxygen, he calculates T_c in the range 30–40 K for doped La_2CuO_4 . An essential part of the argument is that the breathing mode is soft (has low energy) in doped La_2CuO_4 and that it 'freezes' into a lattice distortion in the undoped material. So far so good. But the model also gives T_c , if anything, lower for $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, a 90-K superconductor.

Moreover, recent neutron scattering experiments at Brookhaven's High Flux Beam Reactor⁸ have failed to find signs of

a soft or frozen breathing mode in a single-crystal sample where it surely should have been seen (and where soft modes related to the observed orthorhombic–tetragonal transition were seen). It is difficult to escape the conclusion that something is missing, something that prevents the softening of the breathing mode and removes the best hope for an explanation based on phonons. As we shall see, there are strong antiferromagnetic correlations in the CuO_2 planes (extended antiparallel alignment of electron spins), which do not show up in the band structures⁹ used by Weber, but have a significant influence on the behaviour of the system.

Isotope effect

But first, what of the isotope effect, the original (if fickle) signature that the lattice is involved? Substitution of ¹⁸O for ¹⁶O has a clear effect on T_c in doped La_2CuO_4 (refs 10–12). This suggests that phonon modes involving oxygen are somehow implicated in the superconductivity but it does not prove that they are dominant. Indeed simple estimates, assuming some other mechanism to be the prime mover, show that a relatively small contribution from phonons, perhaps enough to give a T_c of a few degrees, is quite sufficient to produce an isotope effect of the observed magnitude. The estimates indicate a small isotope effect in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ consistent with the null effect (within experimental error) that has been observed.

The isotope effect could be related to one important property of high T_c superconductors — s -state (zero angular momentum) pairing of the kind usually found in metals. This is revealed by the London magnetic penetration depth, as measured by muon spin relaxation^{13,14}. For $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, several experiments show that this quantity follows the usual BCS behaviour (one experiment disagrees¹⁵); the data for doped La_2CuO_4 indicate the same thing, but less clearly because of disorder. The implication is that there are no points or lines on the Fermi surface where the energy gap vanishes, otherwise the temperature dependence would be quite different. Zeros of the gap are an indication of pairing with finite angular momentum, that could help the electrons to avoid a short-range repulsion and to take advantage of a long-range attractive interaction. Absence of zeros is suggestive of s -states but is not fully conclusive. But the phonon-induced interaction is usually isotropic and the only large component has zero angular momentum. Thus, it is natural to conclude that the isotope effect is a symptom of s -state pairing.

A central feature of several purely electronic approaches to high- T_c superconductivity^{16–22} is a strong repulsion between holes in the copper $3d$ -states, leading to the prediction of antiferromagnetic

correlations between spins. The observation of three-dimensional antiferromagnetic order supports this picture, and recent experiments at Brookhaven²³ confirm the existence of two-dimensional antiferromagnetic correlations in La_2CuO_4 extending over very long distances (greater than 50 lattice spacings at low temperatures).

The observations require a very substantial moment on the copper atoms which in turn indicates a strong, short-range, repulsive interaction between electrons. This has important implications for pairing: in particular the correlated motion of the electrons inhibits the softening of the oxygen breathing modes. It also follows that pure La_2CuO_4 is an insulator because of a Mott-Hubbard gap in its energy spectrum. The gap essentially 'freezes out' the electrons already present in La_2CuO_4 so that, on doping, the density of carriers n_c is equal to the density of dopants. This conclusion, and the analogous one for the 90-K superconductors, is quite different from the predictions of electronic structure calculations⁹ but is fully supported by measurements of the Hall effect^{4,6}.

This discussion makes it clear that the oxide superconductors are even more remarkable than they first appeared: they achieve a high T_c despite a strong, short-range, repulsive interaction between electrons and they do so without taking advantage of high-angular-momentum pairing or retardation. Many of these properties, and particularly the low carrier density, appear to be intimately related to antiferromagnetic correlations, which must therefore be built into any complete theory of high- T_c superconductivity. Thus, it is conceptually economical if those correlations also provide the mechanism for superconductivity itself, although so far there is no experimental proof that this is the case.

Typically, such theories are concerned entirely with the quasi-two-dimensional motion of electrons in the CuO_2 planes, in which the copper atoms form a square lattice with oxygen sitting on the bridge sites between them. The valence levels contribute one state per atom, giving a total of three states for each CuO_2 unit cell. Allowing for spin, these states can contain a maximum of six electrons per unit cell: the superconductors have somewhat less than five. One approach¹⁶⁻²⁰ is to identify the states that are partially filled, and to regard them as the active states in the model. Two electrons rarely occupy the same state because there is a strong repulsive interaction between them and the task is to explain superconductivity in the crowded conditions prevailing in the oxides — there are very few vacant sites.

It is customary to carry out a formal rearrangement of the problem by forbidding two electrons of opposite spin to occupy the same state. (The exclusion

principle already does the same thing for parallel spins.) This constraint is excessively costly in kinetic energy and, to compensate, there is an effective attractive interaction between electrons of opposite spin in neighbouring cells. This interaction is held to be responsible for the antiferromagnetic or the resonating-valence-bond¹⁶ correlations, or for superconductivity, depending on the number of vacant states. This mechanism of superconductivity is unusual in that there is no polarization of a medium and nothing is exchanged. One version of this approach was described in the News and Views article by Anderson and Abrahams¹⁶. Other versions differ in important respects (particularly ref. 18).

I have argued for a rather different view of the problem²¹, that retains all three states in the unit cell and first deals with the largest energy in the problem, the interaction between two electrons on copper. It is found that there is essentially one hole (or missing electron) on each copper atom and that the superconductivity is produced by extra holes that are found on the oxygen. An effective attraction is produced by polarizing the spins of the copper atoms. It is strong enough to give a high T_c because of the intimate mixing of copper and oxygen atoms in the particular geometry of the CuO_2 planes. The model thus accounts for the low carrier density and avoids the problem of short-range repulsion, which is much smaller on oxygen

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Cosmic Song, a dynamic sculpture by Serge Moro, recently installed at CERN. Cosmic rays trigger the fluorescent tubes that illuminate the abstract design. (Courtesy of CERN Photo.) □

than on copper. This picture is supported by many X-ray-spectroscopy experiments (see, for example, ref. 24).

We now await more complete experimental data and theoretical calculations to decide on which is the best picture. This description has been necessarily cursory, and I apologize to those whose ideas I have not included. □

1. Bardeen, J., Cooper, L.N. & Schrieffer, J.R. *Phys. Rev.* **108**, 1175 (1957).
2. Esteve, D. *et al. Europhys. Lett.* **3**, 1237 (1987).
3. Gough, C.E. *et al. Nature* **326**, 855 (1987).
4. Ong, N.P. *et al. Proc. Int. Workshop on Mechanisms of Superconductivity*, Berkeley (in the press).
5. Uchida, S. *et al. Proc. Int. Workshop on Mechanisms of Superconductivity*, Berkeley (in the press).
6. Penney, T. *et al. Proc. Int. Workshop on Mechanisms of Superconductivity*, Berkeley (in the press).
7. Weber, W. *Phys. Rev. Lett.* **58**, 1371 (1987); erratum **58**, 2154 (1987).
8. Birgeneau *et al. Phys. Rev. Lett.* (submitted).
9. Mattheiss, L.F. *Phys. Rev. Lett.* **58**, 1028 (1987).
10. Batlogg, B. *et al. Phys. Rev. Lett.* **58**, 2333 (1987).
11. Bourne, L.C. *et al. Phys. Rev. Lett.* **58**, 2337 (1987).
12. Cohen, M.L. *et al. Proc. Int. Workshop on Mechanisms of Superconductivity*, Berkeley (in the press).
13. Aeppli, G. *et al. Phys. Rev.* (submitted).
14. Uemura, T. *et al. Phys. Rev.* (submitted).
15. Ott, H. *Proc. Adriatic Conf. High- T_c Superconductivity*, Trieste (in the press).
16. Anderson, P.W. & Abrahams, E. *Nature* **327**, 363 (1987).
17. Ruckenstein, A.E., Hirschfeld, P.J. & Appel, J. *Phys. Rev. B* **36**, 857 (1987).
18. Kivelson, S.A., Rokhsar, D.S. & Sethna, J.P. *Phys. Rev. B* **35**, 8865 (1987).
19. Schulz, H.J. *J. Phys.* (in the press).
20. Dzyaloshinsky, I. *Proc. Adriatic Conf. High- T_c Superconductivity*, Trieste (in the press).
21. Emery, V.J. *Phys. Rev. Lett.* **58**, 2794 (1987).
22. Hirsch, J.E. *Phys. Rev. Lett.* **59**, 228 (1987).
23. Shirane, A. *et al. Phys. Rev. Lett.* (submitted).
24. Tranquada, J.M. *Phys. Rev. B* **35**, 7187 (1987).

V.J. Emery is at Brookhaven National Laboratory, New York 11973, USA.

Genetic engineering

Satellite defences for plants

Giles Courtice

SATELLITE RNAs are species of RNA associated with some strains of certain plant RNA viruses, but not necessary for virus replication. Satellites are replicated in cells infected with the particular virus they depend on, and because they are packaged with viral genomic RNAs in the virus particle, they can accompany viruses released from diseased plant tissue to new infection sites. Although they are completely dependent on the virus for replication and transmission, their nucleotide sequence seems to be essentially unrelated to that of the viral genome. (Plant virus genomes often consist of more than one RNA species.) Why natural selection does not lead to their loss or elimination is not known, but it is clear that certain satellites reduce, sometimes markedly, the severity of disease symptoms resulting from virus infections where they are present. On pages 799 and 802 of this issue^{1,2}, respectively, Harrison *et al.* and Gerlach *et al.* report that plants containing a DNA fragment in their genome which, when transcribed, gives rise to an RNA satellite of the type that in the wild attenuates disease symptoms, are protected from the effects of viral infection. The practical benefits of this type of work were among the topics discussed at a recent meeting*.

The two groups have worked with different viruses: Harrison *et al.*¹ used cucumber mosaic virus (CMV) and one of its satellites. It was established last year³ that normal satellite RNAs are produced in transgenic tobacco plants containing DNA versions of the satellite. But the only satellite-free laboratory strain of CMV available in the earlier work³ was one that caused only mild symptoms: the new results of Harrison *et al.*¹, obtained with a more aggressive satellite-free strain isolated from the field (most field strains are satellite-free) show that CMV replication and disease symptoms in the satellite-producing transgenic plants are greatly attenuated. The plants are also protected against a related virus, tomato aspermy virus, although with this second virus replication is not greatly reduced — suggesting that protection by the satellite is possible without a reduction in overall virus multiplication.

Gerlach *et al.*², on the other hand, carried out similar experiments with tobacco ringspot virus (TobRV) and its satellite (STobRV). STobRV has attracted attention for its other remarkable properties, but its relevance here is that it

too mitigates the effects of virus infection. Tobacco plants with DNA versions of STobRV in their genome are protected against the deleterious effects of subsequent TobRV infection (see figure).

Both groups find that the best results are obtained when the DNA fragment put into the transgenic plant carries more than a single copy of the DNA version of the satellite. Harrison *et al.* made the CMV satellite in the experimental plants from constructs with 1.3 or 2.3 satellite units³; Gerlach *et al.* transcribed STobRV from triple repeats of permuted monomer of the normal sequence. STobRV can undergo self-splicing *in vitro*, and Gerlach *et al.* believe that the same process occurs *in vivo* in their transgenic plants to yield single copies of the monomer from the initial trimer transcripts.

As well as the practical applications of this approach, more general aspects of the use of genetic engineering in breeding for better resistance were discussed at the meeting. Another strategy⁴ for protection against viruses uses the fact that in at least five systems, transgenic plants that express genes for viral coat proteins are also protected from attack by the virus the protein belongs to (R.N. Beachy, Washington Univ., St Louis). Excess coat protein,

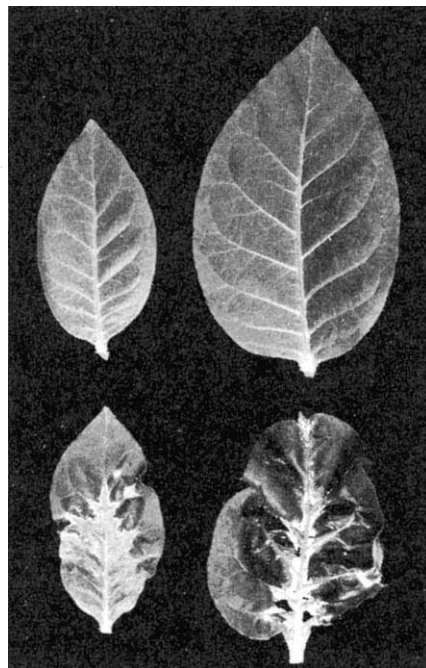
produced by the plant before infection occurs, seems to act by interfering with unwrapping of the plant RNA — the first step in invasion of the plant cell — because naked viral RNA is still infectious to the protected plants. The method does not work with all viral proteins: transgenic tobacco plants expressing the gene for the non-coat '30K' protein of tobacco mosaic virus (TMV), which is believed to be crucial in the spread of virus from an infected cell to neighbouring cells, are only minimally protected (Beachy).

There is a potential problem with the coat-protein strategy. To be sure that reasonable levels of coat protein are already available in case of viral challenge, the transgenic plants must be made to produce the proteins all the time. Such 'constitutive' expression could have effects on the nutritional value of plants protected in this way, and the very difficulty of estimating what these effects might be, or of showing that they do not exist, could delay approval for use in food products from suspicious regulatory authorities.

A feature of the use of satellites in transgenic plants is that production of satellite RNA is switched on by the presence of attacking virus — unchallenged transgenic plants contain only small amounts of satellite. This induction of the response by viral challenge could explain why protection by satellite-gene expression apparently cannot be swamped by inoculation of excessive virus, unlike the coat-protein system, where it seems that excess virus can 'mop up' all the coat protein available, and overcome the protection. This problem might be solved by more sophisticated constructs in which coat-protein genes are under the control of plant promoters where transcription is turned on by pathogen attack.

But there are conceivable plant health hazards in adding to the plant genome sequences encoding satellite RNAs: for one thing, it might allow mutated versions of protective satellites to be produced, for example by the insertion of transposable elements from the plant, and stably inherited. Some natural satellites are known to make disease symptoms worse, and mutants could turn out to be of this sort. Such hypervirulent satellites could be transferred to any strain of the virus that infected a plant containing them, and in this way confer a new aggressiveness on many other strains. Development of disabled, non-transmissible, but still protective satellites might pre-empt this worry.

The satellite-RNA (D. C. Baulcombe, Plant Breeding Institute, Cambridge) and the viral coat-protein methods both seem to provide protection against viruses closely related to the initial target virus, but it is likely that separate gene constructs will be needed for each small group of viruses. As molecular methods reveal



Nicotiana tabacum leaves, 3 weeks after inoculation of the plants with tobacco ringspot virus. Top, leaves containing the synthetic disease-resistance gene constructed using satellite of tobacco ringspot virus. Below, 'normal' leaves, showing effect of virus on untransformed plants. (Courtesy of W. L. Gerlach.)

* Plant Resistance to Viruses CIBA Foundation Symposium 133, London, 31 March–2 April 1987.

more about the biology that underlies these two techniques — in one case the details of virus replication and spread, in the other the life history of the enigmatic satellite RNAs — more sophisticated variants of these strategies will no doubt become possible.

Another possibility discussed at the meeting is that virus replication may be inhibited if the plant is induced to make so-called anti-sense RNA — stretches of RNA complementary to crucial sequences of the plant virus genomic RNA. By forming base pairs to the genomic RNA, such wrecking stretches of RNA might prevent

expression of the virus genome. Initial experimental tests of this strategy give equivocal results (Baulcombe; J. van Emmeloe, Plant Genetic Systems, Gent). The protection obtained is variable and seems to depend crucially on the viral genomic sequences used for blocking. □

1. Harrison, B.D., Mayo, M.A. & Baulcombe, D.C. *Nature* **328**, 799–802 (1987).
2. Gertlach, W.L., Llewellyn, D. & Haseloff, J. *Nature* **328**, 802–805 (1987).
3. Baulcombe, D.C., Saunders, G.R., Bevan, M.W., Mayo, M.A. & Harrison, B.D. *Nature* **321**, 446–449 (1986).
4. Abel, P.P. *et al. Science* **232**, 738–743 (1986).

Giles Courtice is an Assistant Editor of Nature.

Computer vision

Real-time seeing machines?

Andrew Blake

SINCE research into machine vision began, analysis of a single, still image has been quite enough to keep a computer busy for many minutes. Now the promise of ever-increasing computer power, especially with the coming parallel computers, encourages the belief that the rapid analysis of time-sequences of images may soon be practicable. This belief has been accompanied by a surge of research activity in the interpretation of image sequences. A persistent message at a recent meeting* was that, despite their complexity, the inherent richness of information in time-varying images can make them easier to analyse than single, static ones.

This principle is sufficiently pervasive to merit a name — active vision (J. Aloimonos and I. Weiss, University of Maryland; A. Bandyopadhyay, State University of New York) — signifying that an object can spring into three-dimensional relief when a viewer is free to change position. The shapes of textured objects, for example, which might be obscure in a single view, can be interpreted unambiguously from a sequence of images taken as the viewer moves relative to the objects. The viewer's motion has the effect of breaking the camouflage of the object. The idea that motion enhances three-dimensional perception is, of course, familiar. The important advance is that this intuitive notion is now underpinned extensively by mathematical analysis (A. Waxman, B. Kangar-Parsi and M. Subbarao, University of Maryland). It can also be related to properties of human vision^{1,2} by combining experimental observation, mathematical modelling and computational verification.

Recent work (D. Heeger, University of Pennsylvania) shows how a model consisting of a bank of spatio-temporal filters can both predict certain motion illusions and generate useful, accurate estimates of

velocity. As another example of active vision, consider the black silhouette of a vase against a white background. It might have been generated by a real three-dimensional vase out in the world, or by a flat, two-dimensional, cardboard cutout. It is only as a moving viewer observes the unchanging shape of the silhouette that he can conclude it is a real, rotationally symmetrical vase. Recent research (P. Gibling, University of Liverpool; R. Weiss, University of Massachusetts, Amherst) shows how the apparent movement of a silhouette might be used, in quite a general way, to infer object shape.

As a sequence of images evolves with time, it may actually become easier and easier to keep track of objects in the world (N. Ayache and O. Faugeras, INRIA, France). Inferences made by painful analysis of the first few images can be used predictively to ease the task of analysing

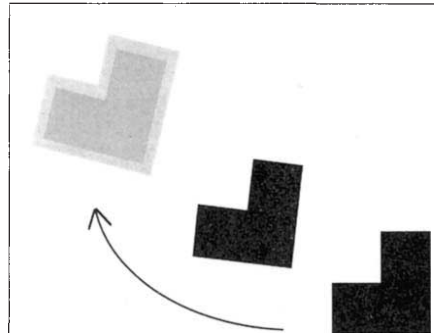


Fig. 1 A moving object is observed at two instants, separated by a short time interval. Estimating the object's motion allows a rough prediction of its position at a third instant.

subsequent images (Fig. 1). In fact, the need to track features in moving images is so fundamental as to call for 'sub-conscious' computational support. A recent provision of this kind maintains rectangular windows, each of which, once placed around a bright moving object, continues to track it automatically. A computer program analysing the moving image is then free to switch attention between windows. A new and elegant approach to the tracking problem is to replace the rigid rectangular windows by flexible 'snakes' (M. Kass, A. Witkin and D. Terzopoulos, Schlumberger, Palo Alto). These snakes can conform precisely to the feature being tracked. This sinuous synthesis of elasticity theory, arcade game and detailed numerical computation has a broad spectrum of possible behaviour. A snake can be given varying degrees of affinities for narrow stripes, abrupt contrast changes and line terminations. Once let loose near a suitable feature, the snake clings tight, even under unpredictable motion. It can lock onto a human lip, a boundary between two different textures or even a

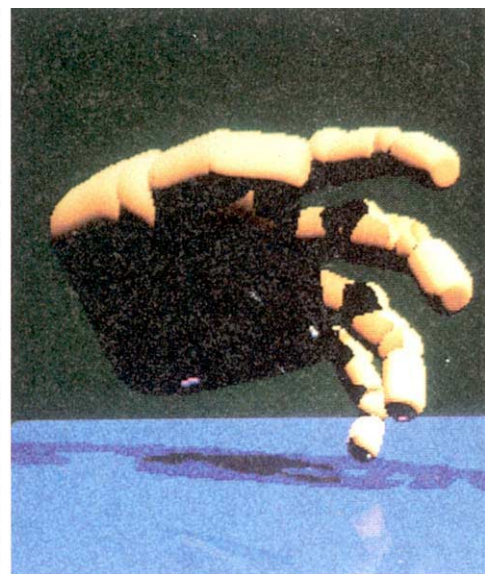


Fig. 2 Superquadrics are sufficiently flexible to model a hand in this computer-aided design system (left). The superquadric description can be used to generate a reasonably realistic shaded image (right). (Images courtesy of I. Elsley and C. Tan, University of Edinburgh.)

* First International Conference on Computer Vision, London, 8–11 June 1987.

'subjective' contour, and track it faithfully.

Many issues remain to be addressed. Although interpretation of various kinds of visual cue may be well understood, the question of integrating disparate sources of information is largely open. Psychophysical studies have shed some light, for instance on combining stereo and shading cues (H. Bülthoff and H. Mallot, Massachusetts Institute of Technology) or stereo and motion (B. Rogers, University of Oxford). In each case, insights come from setting up cues which are mutually contradictory, and seeing which cue dominates. But computer vision systems that have attempted to integrate⁴ information are few and far between.

Another pressing problem is how a three-dimensional object can best be described within a computer program. Simple geometrical entities such as lines and planes have limited representational power. However, they are quite effective in indoor scenes in which most objects have planar surfaces. More general families of shapes include generalized cones⁵ and superquadrics⁶. Superquadrics, for instance, include spheres, ellipsoids, cubes with rounded edges and other shapes. They can be bent, twisted, cut and glued to one another to build up complex

objects. Generalized cones and superquadrics can be used to describe 'natural' objects such as human limbs (Fig. 2) or a subtly shaped artefact such as a vase. How such mathematical descriptions can be turned into images (like Fig. 2) is well understood. The difficult problem — much harder with superquadrics than with lines and planes — is to turn an image into a description. This is the essence of machine vision. In this respect, both generalized cones and superquadrics show signs of possible tractability (R. Horaud, LIFIA; M. Brady, University of Oxford; R. Bajcsy and F. Solina, University of Pennsylvania). Meanwhile, current tried and tested three-dimensional vision systems (O. Faugeras *et al.*, INRIA; J. Mayhew, University of Sheffield) succeed by squeezing the last ounce out of lines and planes. □

1. Ullman, S. *Interpretation of Visual Motion* (MIT, Cambridge, Massachusetts, 1979).
2. Hildreth, E.C. *Artific. Intell.* **23**, 309–354 (1984).
3. Inoue, H. in *Robotics Research* (eds Hanafusa, H. & Inoue, H.) 95–102 (MIT, Cambridge, 1985).
4. Brooks, R.A. *Artific. Intell.* **17**, 285–348 (1981).
5. Marr, D. *Vision* (Freeman, San Francisco, 1982).
6. Barr, A.H. *Computer Graphics* **18**, 21–30 (1984).

Andrew Blake is at the Department of Computer Science, University of Edinburgh, Edinburgh EH9 3JZ, UK.

Chemical physics

Direct probing of reactions

Ian W.M. Smith

THE molecular events that bring about chemical change usually occur with awesome rapidity. When the absorption of a photon breaks a chemical bond or when two species collide and react, the important part of the action is often over in less than a picosecond. As a result, until very recently the choreography of these processes — how the atoms interact and move as a reaction occurs — had to be inferred from the effects of changing the energies or orientation of the reagents, or from observations on the energy states of products or how they are scattered in experiments using molecular beams. Such experiments, which earned Herschbach, Lee and Polanyi the Nobel prize in Chemistry in 1986, concentrate on the 'before' and 'after' of chemical events at the molecular level. Recent experiments using extremely short laser pulses show how it is possible to investigate chemical dynamics actually during the molecular processes that bring about change.

To probe events as short as molecular collisions and dissociations in real time, it is necessary to use the latest short-pulse laser technology. Interestingly, one begins to confront the fundamental limitation imposed by Heisenberg's uncertainty principle. Applied to the output of lasers,

the uncertainty principle means that the shorter a light pulse is, the less well can its frequency or wavelength be defined. With continuous lasers, very high definition of frequency (about 10^{-4} cm⁻¹) is possible, making them today's preferred sources for high-resolution spectroscopy. With a 100-femtosecond (fs) laser pulse, however, the spectral resolution is reduced to about 150 cm⁻¹: the advantage of looking at very short times is being overtaken by the inevitable blurring of the energy resolution.

Laser pulses of 40 femtoseconds have been used in recent work on the photodissociation of ICN. (Dantus, M., Rosker, M.J. & Zewail, A.H. *J. chem. Phys.* **87**, 2395–2397; 1987). Earlier work with 'conventional' 10-ns pulsed lasers had revealed in which states the I atoms and CN radicals are produced if ICN is photolysed with ultraviolet light. But although these measurements can identify the states of the photofragments before they undergo any collisions that modify their states, they are far too slow to observe the species whilst they are still separating from one another. A few years ago, Kinsey and co-workers (Imre, D., Kinsey, D.L., Sinha, A. & Krenas, J. *J. phys. Chem.* **88**, 3956–3964; 1984) realized that a small fraction of photodissociating

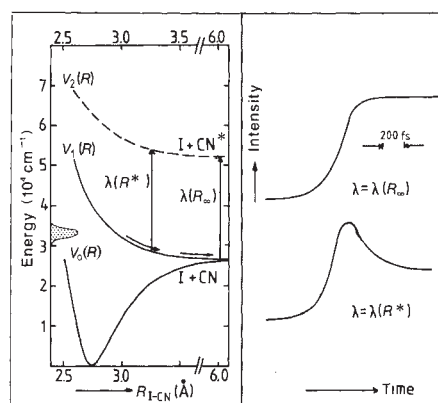


Fig. 1 Femtosecond probing of dissociating ICN. The curves in the panel to the left show the variation of I–CN binding energy V with separation R in three electronic states of ICN. The first (photolysis) laser pulse excites molecules from the ground state, $V_0(R)$ into the dissociating state $V_1(R)$ with an energy distribution represented by the bell-shaped curve. The second (probe) laser pulse can be tuned to different wavelengths λ . With $\lambda = \lambda(R_*)$, signal is only observed once I and CN are well separated so the signal is delayed. With $\lambda = \lambda(R_0)$, the signal rises more rapidly but then falls, as systems pass through the separation R^* for which the energy between V_1 and V_2 matches the photon energy.

molecules (about 1 in 10^6) will fluoresce during the act of separating and that the resultant emission spectrum is a reflection of the dynamics of separation. This ingenious method is indirect, however, and the spectrum is not easily related to the form of the forces acting on the photofragments as they fly apart.

In the experiments of Zewail and colleagues, ICN is prepared with a quite well-defined energy above the dissociation limit of the I–CN bond (see Fig. 1) with one short laser pulse (wavelength 308 nm; duration 100 fs). A second pulse of similar length is tuned to a wavelength at or close to an electronic transition in the isolated CN radical. With the probe laser at the characteristic CN wavelength, fluorescence is excited but there is a measurable delay (about 400 fs) before the signal reaches its maximum value, because the CN has to escape from the I atom before it becomes isolated and its energy levels correspond to the photon energy supplied by the probe laser. On the other hand, if the probe laser is set to a longer wavelength, the signal rises earlier but then falls, as the fragments pass through a separation (say R^* in Fig. 1) which brings the energy levels into coincidence with the photon energy.

These measurements open a new chapter in reaction dynamics where real-time femtosecond spectroscopy will be applied to primary photochemical and chemical processes. The technique particularly lends itself to the study of photoinduced unimolecular reactions. An extra problem arises for bimolecular reactions. Although a short laser pulse can be used to prepare

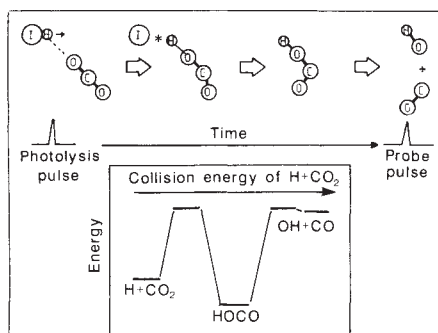


Fig. 2 Picosecond probing of the formation of OH in the reaction of H atoms with CO₂. One picosecond laser pulse breaks the HI bond in a weakly bound HI-OCO species. The delay in the appearance of OH, observed with a second picosecond laser pulse, is caused by the transient formation of an HOCO complex with significant binding energy relative to both H + CO₂ and OH + CO (see boxed insert).

radicals for a rapid bimolecular reaction, they can undergo such a reaction only when they collide. Compared with the timescale of a molecular collision, the time between collisions is usually rather long and there will also be a considerable spread in the times at which prepared radicals react. One way forward is to make use of molecular complexes formed in supersonic molecular beams (Radhakrishnan, G., Buelow, S. & Wittig, C. *J. chem. Phys.* **84**, 727-738; 1985).

In this way, Zewail and co-workers (Scherer, N. F., Khundkar, L. R., Bernstein, R. B. & Zewail, A. H. *J. chem. Phys.* **87**, 1451-1453; 1987) have prepared weakly bound complexes of HI and CO₂. An ultraviolet laser pulse of about 10-ps duration photolyses the HI and impels the H atom towards the CO₂ at considerable velocity (see Fig. 2). Probe laser pulses then measure the appearance of OH from the reaction: H + CO₂ → OH + CO. The short delay of about 5 ps in the appearance of the OH confirms that this reaction, and the exothermic reverse reaction between OH and CO, proceed via the formation of a transitory HOCO collision complex, a proposal that was made, on indirect evidence, about 10 years ago. By initiating the reaction from a bound complex, this technique also defines a preferred orientation for the reagents, an important parameter not usually available to experimentalists.

It seems likely that the techniques of short-pulse, laser technology and molecular beams have now reached a stage where their combination should allow the nature of transition states, even for direct reactions, to be explored for several reacting systems. If transition-state spectroscopy does become established, then the central theory of chemical kinetics will be subjected to exciting new tests. □

Ian W. M. Smith is Professor of Chemistry at the University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.

British IQ

Keeping up with the times

Chris Brand

PERFORMANCE on culture-fair IQ tests is improving worldwide (see the News and Views article last month in *Nature* **328**, 110; 1987). Yet does the average gain conceal important differences between nations? James Flynn, the scholarly reporter from New Zealand of the general trend, seems to think so (*Psychol. Bull.* **101**, 171-191; 1987). He especially remarks not only the paucity of British data on IQ (thanks to years of neglect by Britain's educational establishment) but also a tendency for such reported British gains as he could trace to be modest by the standards of other advanced economies. But, coming to the rescue, Richard Lynn and co-workers on page 797 of this issue provide evidence of a rise in IQ in England averaging 0.25 points annually over the past 50 years.

Lynn *et al.* tested a sample of 1,029 9-11-year-old children from several urban and rural locations in England with the same test that the psychometrician Raymond Cattell had first used in the city of Leicester in 1935 and 1936. Their result is simplicity itself. English children today have a mean IQ of 112 (in comparison with an IQ of 100 in 1935). A minor complication is that the rate of IQ gain may have moderated a little during the Second World War, only to accelerate more steeply in subsequent years. One possible reservation about the study of Lynn *et al.* is that more low-IQ children of today may be withdrawn into special schools; yet any such change should be roughly counterbalanced by the re-routing of higher-IQ children into private schools because of parental anxieties about falling educational standards in state-run schools.

So what is the meaning of the general IQ rise that has now received its English testimony? Importantly, the 'massive' IQ gain — in the study of Lynn *et al.*, just as in Flynn's data from 14 countries — is principally derived from culture-fair, non-verbal IQ tests that require no particular educational achievement. The possibility is that such gains reflect lowered levels of caution, conscientiousness and conservatism of social attitudes — changes which Lynn has himself recorded elsewhere as increasingly 'extroverted' tendencies in the West since the Second World War (Lynn, R. & Hampson, S. L. *Brit. J. soc. clin. Psychol.* **16**, 131-138; 1977). For extroversion (or perhaps 'recklessness', indexed as it was in Lynn and Hampson's study by national levels of divorce, cigarette consumption, accidents, crime and illegitimacy) has been rising much faster than has culture-fair IQ: using

figures from 1950 onwards, Lynn and Hampson's work suggested gains of around 2 standard deviations over a 30-year period in advanced countries. This gain should be contrasted with the rise of merely 0.5 of a standard deviation in culture-fair IQ per generation exhibited by English children. Moreover, extroverted and self-reported delinquent tendencies are well known to be associated with somewhat higher culture-fair IQs (for example, Ley, P., Spelman, M. S., Davies, A. D. M. & Riley, S. *Br. J. educ. Psychol.* **36**, 185-191; 1966) — sometimes correlating as highly as 0.46 with timed performance on such IQ tests (Jensen, A. R. *Co-op Project No. 1867*, US Dept Health, Education & Welfare; 1964). By contrast, vocabulary levels, always the best indicator of general intelligence in normal, non-deprived circumstances, have no such association with behavioural disinhibition; and even Flynn does not claim that there has been any great increase in vocabulary scores.

The question that remains in the aftermath of the present work of Lynn *et al.* is clear. Has the IQ rise occurred mainly on tests of culture-fair IQ that may always have given less conscientious people an unfair advantage? A stress in the classroom on speed of response, imagination and intuition might have improved performance on culture-fair tests, most of which require relatively little by way of articulate judgment or analytical scrupulousity. Certainly, if this were so, it would help explain why conservatism in social attitudes is well known to be negatively correlated with culture-fair IQ (at around -0.40, even when age is controlled). England's culture-fair IQ rise would then be compatible with its even more rapidly increasing rate (until recently) of crime, irreligiosity, sexual promiscuity and anti-authoritarianism; and also with its world leadership in the field of popular music.

Whether Britain will ever again take the lead in the testing of IQ is a different matter. The study of Lynn *et al.* shows what can be done even without the help of state funding; but interest will now focus on whether the Scottish Education Department will resolve to examine the intelligence of Scotland's children today with the same verbal IQ test which was first administered to an entire cohort of Scottish children in 1932 and which showed a mere 2 IQ-point gain when it was re-administered in 1947. □

Chris Brand is in the Department of Psychology, University of Edinburgh, 7 George Square, Edinburgh EH8 9JZ, UK.

Astrophysics

Models of quasars reappraised

Martin Elvis

THE most fully developed method of studying quasars is probably the investigation of their broad emission lines. This area of quasar research is one of the few in which important free parameters are tightly constrained by observations. It became apparent at a recent workshop* that something is terribly wrong with the models built up over the past decade. A result of this crisis could be that quasar emission lines are yet more valuable than had previously been realized.

Quasars emit line radiation from all the common elements over a wide range of ionization levels. These lines are very broad, indicating Doppler velocities of about $10,000 \text{ km s}^{-1}$. To the first order, all the lines have similar smooth profiles, suggesting that they come from an ensemble of small gas clouds moving around the central engine of the quasar. This engine emits a strong non-thermal continuum that ionizes the gas in its vicinity. The gas is held at a temperature of $10\text{--}20,000 \text{ K}$ because any extra heat is efficiently radiated away by emission lines. The density of the gas cannot be more than $3 \times 10^9 \text{ cm}^{-3}$ because we see the C III] (twice-ionized carbon) line at $1,909 \text{ \AA}$, which is quenched at higher densities. On the other hand, the absence of broad [O III] emission implies that the density is at least 10^6 cm^{-3} for the same reason. Elucidating what happens in this situation is a complex, but solvable, problem that depends mainly on the correct use of atomic physics.

Line ratios

There is agreement that models using an ionizing continuum that extends to kiloelectronvolt (keV) X-rays will reproduce the observed line ratios for most of the important lines — C IV, C III], Mg II and the Lyman- α , Balmer- α and Balmer- β lines of hydrogen (Ly α , H α , H β). These models reproduce the low observed values of the Ly α /H β ratio¹ if the cloud is sufficiently thick for ultraviolet rays to be excluded from the interior which is then heated by the X-rays. It is these models, known collectively as the standard model, that are now under attack.

In the standard model, the ionizing continuum also produces an instability in the illuminated region so that two phases of the gas at 10^4 and 10^7 K arise naturally and coexist in pressure equilibrium². This instability arises only for a limited range of the ionization parameter (the ratio of the

photon and electron densities) that is also the range required by the emission-line ratios. This unexpected result gave extra strength to the standard model. Furthermore, this effect prevents the proposed small clouds from dissipating.

S. Collin-Souffrin (Institut d'Astrophysique, Paris) summarized the difficulties of producing the different properties of high- and low-ionization lines using the standard models, which cannot account either for strengths of the O VI and N V lines or for the observed differences in the line profiles and the rapid variability of the lines. A major problem is in reproducing the strong Fe II emission (M. Joly, Meudon) that is observed³. The approximations used in the model could be the cause of this difficulty, which might be resolved if the complete physics of iron were used.

Even the success of the Ly α /H β ratio prediction is now questionable. G. Ferland (Ohio State Univ.) pointed out that the model only gives the observed low values if the X-ray output of the quasar is abnormally strong. The necessary flux of X-rays is seen in 3C273 and a few other objects, but most quasars are ten times too weak. More usual continuum forms produce Ly α /H β ratios that are several times too large. They also have a Compton temperature that is too low to sustain the two-phase instability, so that only a single temperature can be maintained in the gas. Without the instability, the gas clouds will disperse unless other means of confinement exist such as magnetic fields (M. Rees, Cambridge Univ.) or gravity (M. V. Penston, Royal Greenwich Observatory).

As well as theoretical problems, observations also demand a reassessment of the physical state of the gas clouds. The spectrum indicates a well-defined ionization parameter in the gas. Because the density of the gas is known, the distance of the gas from the source needed to produce the correct photon density can be deduced. This distance is about 1 light year for a low-luminosity active galaxy and about 10 light years for brighter quasars. The time variability of several objects suggests that the actual size of the region emitting the broad lines is about an order of magnitude smaller than this. Penston recalled the extensive International Ultraviolet Explorer study of NGC4151 that indicates a size of $10\text{--}20$ light days. D. Alloin (Meudon) and B. Peterson (Ohio State Univ.) presented analyses of data on the Seyfert galaxy Akn 120. Although some disagreement was

expressed at the meeting, there is little doubt that variability occurs on a 30-day timescale in this object, ten times quicker than expected. Several other similar cases were presented.

Even if the timescales we expect are wrong on an absolute scale, we still expect a clear association of longer timescales with higher luminosities, because the cloud systems should be larger in brighter quasars. But P. Gondhalekar (Rutherford Appleton Laboratory) showed examples of quasars that change their luminosity in 100 days or less. Unfortunately, there were typically only two observations per quasar, so that he could not set a timescale. Nevertheless, the fact that a quarter of his objects showed such changes argues that variability in a few tens of days is common in these high-luminosity objects and the timescale does not seem to depend strongly on luminosity. One of the assumptions must be wrong. Either the gas densities are higher and C III] emission comes from somewhere else; the ionization parameter is large, which would make extra difficulties with the line ratios; or there is a special geometry in the nucleus so that, for example, the gas does not see the same continuum that we do.

New approach

Can the standard model be saved or is an entirely new approach needed? Ferland noted some ways in which the standard model is incomplete: it does not consider γ -ray or microwave heating. One model that does not make the usual assumption that photons are created and re-absorbed locally shows non-local effects that can be significant⁴ but, because it uses only pure hydrogen, it is not directly comparable to other work and has been discounted. Recent work⁵ on a similar model that includes all elements shows that these non-local effects can indeed make the difference to the crucial ratios. It is not obvious at this stage, though, in which direction these factors will go in any particular case.

Another approach (J. Perry, Cambridge Univ. and J. Dyson, Manchester Univ.) involves the construction of a complete physical model of the emission-line region of the quasar. This gives a broader perspective and allows a wider range of constraints to be included. Unfortunately this is an all-or-nothing approach that does not allow the problem to be solved a piece at a time. In fact, by the time all the details are included, the model becomes so complex that it cannot be unique.

If the clouds are not photoionized by the quasar continuum then some collisional process could be at work. Collin-Souffrin made a plea for the investigation of the high-velocity shocks that would occur if two clouds collided.

Only a small percentage of the total

* *Emission Lines in Quasars and Active Galaxies*, Rutherford Appleton Laboratories, 27–29 May 1987.

luminosity of a quasar is emitted in the lines. For this reason, some people have decided to ignore them. After all, is it worth expending effort on something so indirectly related to the central engine of the quasar? On the other hand, as the emission-line problem has turned out to be so intractable, researchers are now looking for extra sources of energy in quasars. □

1. Baldwin, J.A. *Mon. Not. R. astr. Soc.* **178**, 67p–74p (1977).
2. Krolik, J.H., McKee, C.F. & Tarter, C.B. *Astrophys. J.* **249**, 422–442 (1981).
3. Wills, B.J., Netzer, H. & Wills D. *Astrophys. J.* **288**, 94–116 (1985).
4. Hubbard, E.N. & Puetter, R.C. *Astrophys. J.* **290**, 394–410 (1985).
5. Arrett, E. *Astrophys. J.* (submitted).

Martin Elvis is at the Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02183, USA.

Signal recognition

Two receptors act sequentially

Peter Walter

SIGNAL sequences specifying the proteins that are targeted to and translocated across the membrane of the endoplasmic reticulum in eukaryotic cells have a wide degree of sequence diversity. The main common denominator appears to be a clustering of hydrophobic amino acids, hardly a feature unique enough to ensure proper intracellular protein localization. Despite this diversity, signal sequences seem to be recognized by specific protein receptors that decipher their information content. Two such receptors are now known. The paper by Wiedmann *et al.* on page 830 of this issue¹ shows that the 'classical' receptor, the signal-recognition particle (SRP), shares its burden with another protein, the signal-sequence receptor (SSR), an integral membrane protein of the endoplasmic reticulum.

According to the latest hypothesis of how the protein-translocation apparatus works, signal sequences are first recognized by the SRP as they emerge from the ribosome (Fig. 1a). (Note that for my subsequent discussion, the SRP has been demonstrated to bind to signal sequences only in the context of the ribosome.) Then the interaction of the tightly bound SRP with the SRP receptor (also called the docking protein), an endoplasmic reticulum membrane protein, provides for the specificity of the targeting reaction of the ribosome/nascent chain/SRP complex to the correct intracellular membrane. There, the ribosome engages in a ribosome-membrane junction, presumably by binding to ribosome receptors, whereas other membrane proteins may assemble a translocation pore. In this process, the signal sequence is handed over to the SSR without the need for further peptide elongation. The SRP and its receptor are released from the membrane-bound ribosome and are presumably free to recycle. How subsequent protein translocation takes place is largely unknown, but we assume — by analogy with protein translocation across mitochondrial membranes — that the protein is in an extended conformation as it passes through the membrane. The signal sequence must dissociate from the SSR

again as the nascent chain elongates further. At this stage the signal sequence may interact with lipids or yet other proteins that are part of the translocation pore. The SSR, which is glycosylated and thus spans the lipid bilayer completely, could be an integral part of this assembly.

The identification of the SSR by Wiedmann *et al.*¹ is based on a very elegant crosslinking approach, a method resulting from the pioneering work of Johnson *et al.*², who showed that the epsilon-amino group of lysine in charged lysyl-transfer RNA can be selectively modified. The altered lysine residue will still be accepted by the ribosome and incorporated into newly synthesized protein. Using this approach to incorporate a photoactivatable crosslinking reagent into the signal peptide of nascent preprolactin, it was established that the protein subunit of the SRP with a relative molecular mass of 54,000 contains the signal-sequence binding site^{3,4}. Extending these studies one step further, Wiedmann *et al.* monitored transfer of the signal sequence from SRP to an integral, glycosylated membrane protein (SSR). The interaction of the signal sequence with SSR is probably a true intermediate step in the translocation reaction, because the signal sequence was presented to SSR in a physiological context (emerging from the ribosome, targeted by SRP/SRP receptor and, most important, if not stopped by the crosslinking, leading to properly translocated product). Thus, this approach seems preferable to the use of isolated signal peptides, which, because of their intrinsically hydrophobic nature, may engage in other, more nonspecific interactions, or may bind only with very low affinity (their binding to isolated SRP, for example, has yet to be demonstrated). However, cross-linked products of membrane proteins to synthetic signal peptides have been obtained⁵ and their relationship to the SSR remains to be determined.

Because crosslinking experiments only demonstrate proximity and do not readily address the specificity of binding events, the identification of the SSR will not be a complete surprise to observers in the field.

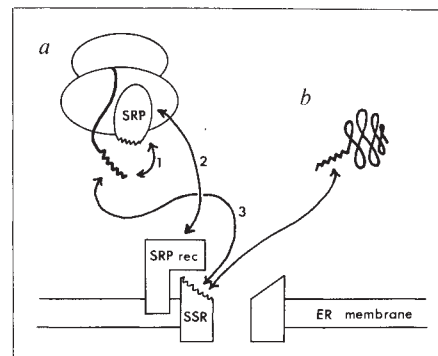
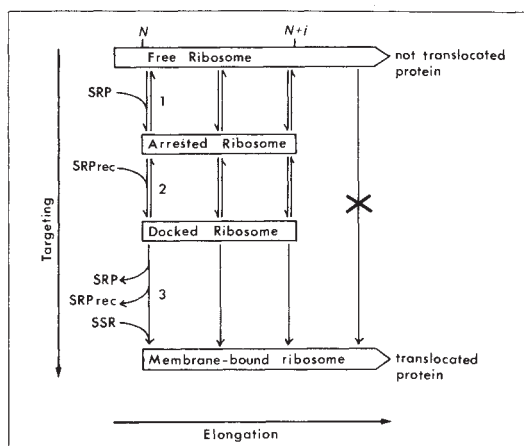


Fig. 1 Information flow during targeting. *a*, Facilitated recognition during co-translational targeting (defined as ribosome/SRP/SRP receptor dependent). The sequential interactions are: 1, signal sequence → SRP; 2, SRP → SRP receptor; and 3, signal sequence → SSR. This mode is used for protein translocation across the mammalian endoplasmic reticulum; SSR is not directly accessible for substrates that have been released from the ribosome (see *b*). The only exceptions are very tiny proteins that may interact with SSR directly. *b*, Direct recognition during post-translational targeting (ribosome/SRP/SRP-receptor independent). The interaction signal sequence → SSR can occur, bypassing the requirement for recognition by SRP on the ribosome. This mode can occur for translocation of at least a few proteins across the yeast endoplasmic reticulum *in vitro* (prepro- α -factor^{17,19}, preprocarboxypeptidase Y (W. Hansen & P.W., unpublished) and prokaryotic plasma membrane.

The existence of a signal sequence receptor in the endoplasmic reticulum membrane was suggested previously from the observations that signal sequences, either as part of small preproteins⁶ or as synthetic peptides^{5,7}, can block protein translocation as competitive inhibitors when pre-incubated with the membranes. In addition, when ribosomes are released from the membrane by puromycin treatment, short nascent chains that have not yet been translocated and cleaved by signal peptidase will remain membrane-bound in what appears operationally as a protein/protein interaction⁸.

In yeast, the translocation of various secretory proteins across the endoplasmic reticulum membrane can occur post-translationally, that is, independent of ribosome and SRP (see the recent discussion in News and Views⁹). Although it is assumed that a yeast SRP analogue exists (it has, however, not been positively identified), to the best of our knowledge no such particle is involved in the post-translational translocation reaction. In Fig. 1, I show a unifying model in which in yeast endoplasmic reticulum membranes an SSR analogue is available for a direct interaction with signal sequences on completely translated proteins which can then enter the translocation pathway (Fig. 1b). It is not yet clear to what degree direct targeting to the SSR bypasses another postulated SRP-dependent targeting

Fig. 2 Kinetic model for SRP-dependent targeting (adapted from ref. 12). 1 (compare with step 1 in Fig. 1a). SRP will interact with a ribosome and arrest protein elongation, after N amino acids have been polymerized and the signal sequence is exposed outside the ribosome. This interaction will only occur within a window of nascent chain lengths (N to $N+i$). Once elongation proceeds beyond this window, the affinity of SRP for the ribosome/nascent chain complex drops drastically, presumably because of aberrant folding that renders the signal sequence inaccessible. The size of the window can vary widely between different proteins depending on their folding characteristics. Some proteins can be synthesized to full length and still be translocated across mammalian membranes provided that the protein has not been released from the ribosome, that is, terminated^{20,21}. Outside the window, no targeting or translocation will take place. Importantly, the model treats the binding of SRP as an equilibrium reaction: the affinity constant will determine how far protein synthesis is retarded in a membrane-free system, where the concentration of SRP receptor is effectively zero. The affinity constant depends on the individual signal sequence, possibly its position in the protein and the nature of the ribosomes. 2 (see step 2 in Fig. 1a). The interaction of the bound SRP with SRP receptor leads to a correctly targeted, docked ribosome. Again, the interaction is, in principle, reversible. 3 (see step 3 and beyond in Fig. 1a). SRP and SRP receptor leave and are presumably recycled. Elongation proceeds. The signal sequence binds to SSR. A functional ribosome-membrane junction is established that allows for translocation of the nascent polypeptide chain.



reaction or whether it might be a special feature used exclusively by certain proteins.

It may even be possible to trace a functional SSR as far back as prokaryotic cells. A candidate for SSR in *Escherichia coli* might be the *prlA* (or *secY*) gene product, an inner membrane protein essential for protein secretion. This assignment is based on genetic evidence suggesting a direct interaction of this membrane protein with signal sequences. Mutants in *prlA* are capable of suppressing various signal sequence defects¹⁰, that is, by altering the *PrlA* protein the signal-recognition machinery is rendered more 'sloppy' and accepts sequences it would normally discriminate against.

Given that in the endoplasmic reticulum membrane a SSR exists which, by definition, recognizes these sequences during the normal translocation process, the question arises of why a SRP would be needed at all. In effect, SRP can be thought of as a soluble signal-sequence receptor whose action precedes that of the SSR and at first glance appears to be functionally redundant. The teleological rationale for this apparent redundancy is not clear, but there are some attractive and not necessarily mutually exclusive possibilities. First, a dual recognition by two receptors functioning sequentially may allow higher specificity in the recognition event. If binding of signal sequences to SRP and SSR is distinguished by emphasizing different subtleties in their structure, then the requirements for SRP-SSR-dependent translocation may be more stringent than if only SSR were used. Here, SRP would increase the fidelity of the targeting event.

Second, the SRP can be thought of as the adapter molecule between the cytoplasmic translation and the membrane-bound protein translocation machineries rather than 'just' as a signal-sequence receptor. As such, the function of the SRP should be regarded as closely intertwined with translation. The fact that it contains an RNA molecule already implies that the particle evolved to interact with other RNA machines. It appears a basic prerequisite for protein translocation across the mammalian endoplasmic reticulum membrane that the substrate protein is targeted while emerging from the ribosome. Even prepro- α -factor, a yeast secretory protein that traverses the yeast endoplasmic reticulum very efficiently post-translationally, will not do so across mammalian endoplasmic reticulum (see Fig. 1a). Thus, unlike yeast, mammalian cells have become addicted to SRP, in that the establishment of a ribosome-membrane junction through the SRP/SRP receptor is essential and interaction of signal sequences with SSR directly either cannot occur or does not suffice for subsequent translocation.

Supporting this functional importance of the SRP/ribosome interaction is the ability of the SRP to slow the rate of protein elongation after binding to a signal peptide as part of the nascent chain. This activity is contained in a separable domain of SRP and is not required for the basic signal recognition and targeting function of SRP¹¹. According to a recently proposed kinetic model¹² (Fig. 2), the elongation-arrest activity increases the length of time during which a nascent protein remains translocation-competent. As elongation is retarded, so are the chances

of the nascent protein to fold — either aberrantly and thereby making the signal sequences unavailable for subsequent interactions, or so tightly that the tertiary structure will impede subsequent translocation. The overall effect is to render protein translocation more efficient.

Finally, George Orwell's rule that "whatever is not forbidden, is compulsory" may be generally applicable in nature. Thus, given that SRP can modulate translation, this activity could provide a regulatory mechanism by which cells could adapt translation for their secretory needs. Indeed, preliminary evidence suggests¹³ that insulin synthesis is regulated at the translational level by a mechanism that involves SRP.

Thus, taken together, SSR may be the evolutionary progenitor of signal recognition at the membrane, where protein translocation is to take place. SRP may have evolved to allow for a tighter coupling between translation and translocation, possibly to allow the superimposition of additional control mechanisms. By using almost exclusively a co-translational translocation mechanism, mammalian cells have been relieved of the need to pay attention to the particular structural nature of their secretory proteins: a polypeptide chain that is translocated as it is being made has no chance to fold in the cytoplasm and therefore no unfolding needs to be considered¹⁴.

Two mammalian proteins have now been described that interact with signal sequences. Precisely what features in signal sequences SRP or SSR recognize to provide specificity to the binding reaction remains a mystery which will be solved only by a high-resolution structural analysis. □

1. Wiedmann, M., Kurzchalia, T., Hartmann, E. & Rapoport, T.A. *Nature* **328**, 830-833 (1987).
2. Johnson, A.E., Woodward, W.R., Herbert, E. & Menninger, J.R. *Biochemistry* **15**, 569 (1976).
3. Kurzchalia, T.V. *et al.* *Nature* **320**, 634 (1986).
4. Krieg, U., Walter, P. & Johnson, A.E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8604 (1986).
5. Robinson, A., Kaderbhai, M.A. & Austen, B. *Biochem. J.* **242**, 767 (1987).
6. Prehn, S., Nuernberg, P. & Rapoport, T.A. *FEBS. Lett.* **123**, 79 (1980).
7. Majzoub, J.A. *et al.* *J. biol. Chem.* **255**, 11478 (1980).
8. Gilmore, R. & Blobel, G. *Cell* **35**, 677 (1983).
9. Schatz, G. *Nature* **321**, 108 (1986).
10. Emr, S.D., Hanley-Was, S. & Silhavy, T.J. *Cell* **23**, 79 (1981).
11. Siegel, V. & Walter, P. *Nature* **320**, 81 (1986).
12. Rapoport, T.A., Heinrich, R., Walter, P. & Schulmeister, T. *J. molec. Biol.* **195**, 621 (1987).
13. Welsh, M., Scherberg, N., Gilmore, R. & Steiner, D.F. *Biochem. J.* **235**, 459 (1987).
14. Rothman, J.E. & Kornberg, R.D. *Nature* **322**, 209 (1986).
15. Schlensted, G. & Zimmermann, R. *EMBO J.* **6**, 699 (1987).
16. Perara, E., Rothman, R.E. & Lingappa, V.R. *Science* **232**, 348 (1986).
17. Hansen, W., Garcia, P.D. & Walter, P. *Cell* **45**, 397 (1986).
18. Waters, M.G. & Blobel, G. *Cell Biol.* **102**, 1543 (1986).
19. Rothblatt, J. & Meyer, D.I. *EMBO J.* **4**, 1031 (1986).
20. Caulfield, M.P., Duong, L.T. & Rosenblatt, M. *J. biol. Chem.* **261**, 10953 (1986).
21. Mueckler, M.K. & Lodish, H. *Nature* **322**, 549 (1986).

Peter Walter is in the Department of Biochemistry and Biophysics, University of California Medical School, San Francisco, California 94143-0448, USA.

Observational constraints on interstellar diamonds

SIR—In a recent study Lewis *et al.*¹ discovered in meteoritic material small (50 Å) diamond grains of extrasolar origin. They were apparently unaware of the previous work² which had proposed that interstellar dust contains diamonds. Adequate observations are now available to test this hypothesis.

Diamond has an absorption edge in the ultraviolet³ beginning at ~ 7 eV with an absorption coefficient, α , of $5 \times 10^5 \text{ cm}^{-1}$ at 7.5 eV. Following earlier work⁴⁻⁶, the optical depth, in magnitudes, can be expressed as

$$\tau(\lambda) = 1.1 C_{\text{abs}} NL \quad (1)$$

where C_{abs} is the absorption cross-section for a small spherical grain, N is the number of grains per cm^3 and L is the path length in cm. The quantity NL can be expressed as

$$NL = \frac{N_{\text{H}} E(B-V)}{6 \times 10^{23}} \frac{W_{\text{C}}}{n_{\text{c}}} \frac{A_{\text{c}}}{\left(\frac{4}{3} \pi a^3\right)} f_{\text{c}} \quad (2)$$

where N_{H} is the number of hydrogens per cm^2 -magnitude, $E(B-V)$ is the reddening, W_{C} is the atomic weight of carbon, n_{c} is the density of the condensed diamond grain, a is the grain radius, A_{c} is the relative abundance of available carbon, and f_{c} is the fraction of available carbon condensed in grains.

To evaluate equation (1) we have used appropriate values for N_{H} (5.8×10^{21} , ref. 7), A_{c} (4.8×10^{-4} , ref. 6), α (5×10^5 at ~ 7.5 eV, ref. 3), n_{c} (3.5, ref. 7), and ϵ , the dielectric constant of diamond³ ($11 + 4i$ at ~ 7.5 eV). The value for α is taken³ as 0 below ~ 7 eV. Using these values and appropriate expressions^{3,8} for α and C_{abs}

$$\frac{\tau(7.5 \text{ eV}) - \tau(7 \text{ eV})}{E(B-V)} = \frac{\tau_{\text{acc}}}{E(B-V)} = 1.6 f_{\text{c}} \quad (3)$$

where τ_{acc} is the differential absorption for the diamond absorption edge. This expression is independent of particle radius for $a < 100 \text{ Å}$.

Because the absorption edge, which continues to increase with increasing energy, occurs in the same energy region as the proposed crystalline MgO feature⁵ we can apply the recent results derived by Massa *et al.*⁹. They found no fine structure in the interstellar extinction curve between 3,000 and 1,200 Å. Using this and equation (3) puts an upper limit of 1% for the amount of interstellar carbon condensed as small crystalline diamond grains. While this upper limit does not rule out the extrasolar origin of the meteoritic diamonds, as only 1% of the meteoritic carbonaceous material was in the form of small diamonds, it should be compared with the results of recent models that show 60–80% of carbon condensed in grains^{6,8}. Thus, if small diamonds exist in inter-

stellar dust they are a minor species.

Laboratory studies may provide a possible solution to the origin of the meteoritic diamonds. These studies¹⁰ show that condensation from a vapour containing C and H results in an amorphous material which has properties and bonding that are characteristic of both diamond (sp^3) and graphite (sp^2). The proportion of the two types of bonding varies, and heating of this amorphous CH material can cause graphitization. Based partly on these results, Hecht⁸ proposed a model for interstellar and circumstellar dust where nearly 20% of C was in the condensed form, as either small amorphous CH grains or small partially graphitized grains formed from heating the original amorphous material. A possible explanation for the meteoritic diamond grains consistent with this model would be as follows. Some of the small amorphous CH grains may have a large excess of diamond bonds. Subsequent shock heating or particle bombardment would result in partial crystallization in a diamond form. This would explain the observed character of the meteoritic diamond¹: that is, a diffraction pattern indicating the presence of crystalline diamond, a low density indicating the presence of amorphous material, infrared features indicating the presence of soot, and a varying colour of the bulk material. The phase change from amorphous CH to crystalline diamond probably does not occur in circumstellar shells, as no infrared features² associated with diamond have yet been seen. Because the diamonds already show evidence of supernova interaction, it may be that the crystallization occurred as the grains passed by a supernova and the isotope enrichment occurred after this. An alternative explanation is that formation actually occurred in the ejecta from a supernova. Thus, evidence for the grains such as increased ultraviolet extinction near 7 eV or infrared features² from 4–5 μm and 8–9 μm should be looked for in the current supernova. This work was supported by the Aerospace Sponsored Research Program.

J. H. HECHT

Space Sciences Laboratory,
The Aerospace Corporation,
Los Angeles, California 90009, USA

1. Lewis, R.S., Ming, T., Wachter, J.F., Anders, E. & Steel, E. *Nature* **326**, 160–162 (1987).
2. Saslaw, W.C. & Gaustad, J.E. *Science* **221**, 160–162 (1969).
3. Philipp, H.R. & Taft, E.A. *Phys. Rev.* **127**, 159–161 (1962).
4. Van de Hulst, H.C. *Light Scattering by Small Particles* (Dover, New York, 1981).
5. MacLean, S. & Duley, W.W. *Astrophys. J. Lett.* **252**, L25–L27 (1982).
6. Draine, B.T. & Lee, H.M. *Astrophys. J.* **285**, 89–108 (1984).
7. Bohlin, R.C., Savage, B.D. & Drake, J.F. *Astrophys. J.* **244**, 132–142 (1978).
8. Hecht, J.H. *Astrophys. J.* **305**, 817–822 (1986).
9. Massa, D., Savage, B.D. & Fitzpatrick, E.L. *Astrophys. J.* **266**, 662–683 (1983).
10. Fink, J. *et al. Phys. Rev.* **B30**, 4713–4718 (1984).

Causal factors in generational IQ gains

SIR — I enjoyed Chris Brand's News and Views article¹ that discusses my work² showing that young adults in 1980 out-scored young adults in 1950 by 18 points, at least on culturally-reduced IQ tests. However, Brand assumes that he and I disagree about the implications of massive IQ gains more than we actually do.

The fact that IQ tests cannot compare successive generations for intelligence does not necessarily mean that they cannot measure intelligence within a particular generation. My best guess is that the tests cannot compare nations, say the United States and Japan, with certain ethnic groups, say American whites and American Orientals. They can certainly compare the gifted and the retarded and may work reasonably well between social classes. Presumably it is the cultural distance travelled from one generation to another that increases IQ scores without increasing intelligence.

Only when we know what casual factors are at work will we be able to assess the significance of the cultural distance between other groups.

The importance of the causal question prompts a few comments on factors Brand suggests. The gains are too great for test sophistication to be significant. Administering and scoring Raven's is virtually automatic, which excludes tester laxity and skills; Brand's table casts doubt on nutrition as the Dutch 18-year olds tested in 1962 were born during the great Dutch famine of 1944 and yet introduce no incongruity in the pattern of gains; increased outbreeding is supposed to have occurred far more before 1950 than after; and family size has both increased and decreased during the period of massive gains. Flieller *et al.*³ give data showing that urbanization is responsible for only three points of a 24-point French gain. The number of years of schooling is a negligible factor at least in Holland³. The qualitative changes in schooling Brand describes might explain gains on some IQ tests but would dictate losses on others, which have not occurred.

Massive IQ gains are truly baffling: it is as if we had suddenly discovered a dramatic but unexplained escalation in juggling skills, and no one could find a carry-over to socially significant sports such as soccer and cricket.

JAMES R. FLYNN

Department of Political Studies,
University of Otago,
PO Box 56, Dunedin,
New Zealand

1. Brand, C. *Nature* **328**, 110 (1987).
2. Flynn, J.R. *Psychol. Bull.* **101**, 171 (1987).
3. Flieller, A., Saintigny, N. & Schaeffer, R. *L'Orientation Scolaire et Professionnelle* **15**, 71 (1986).

Prokaryotic and eukaryotic cell-cycle proteins

SIR—We report here what we believe to be the first demonstration of homology between prokaryotic and eukaryotic cell-cycle proteins. It is the result of a comparison of the predicted amino-acid sequence of the *Escherichia coli* cell-cycle protein FtsA¹ determined by Robinson *et al.*² with the National Biomedical Research Foundation protein sequence database (Version 8; 809,386 residues) using the 'best local similarity' algorithm of Smith and Waterman³ as implemented on an ICL DAP supercomputer by Lyall *et al.*⁴. A search of this type can give valuable insights into the possible functions and origins of predicted gene products as similarity of sequence suggests similarity of structure and function.

Alignments of amino-acid sequence are selected by the program from all those possible, including insertions and deletions where required in order to improve the quality of the alignment. Insertions and deletions are penalized to be worse than the worst mismatch scored. The analysis revealed a highly significant similarity, extending over a region of 60 amino acids, between the primary sequence of FtsA and the cell-cycle proteins CDC28 from the budding yeast *Saccharomyces cerevisiae*⁵ and CDC2 from the fission yeast *Schizosaccharomyces pombe*⁶. The figure shows the extent of similarity observed between the three proteins.

We used the conserved elements in the region shown as a pattern to re-search the database for related segments and to assess their significance. The alignment of the conserved pattern with CDC28, for example, scored more than 60 standard deviations above the expected frequency, based on statistical analysis of the distribution of the scores of the best alignments. The homology reported between these proteins is therefore extremely significant; the next best alignment scored only 2 standard deviations above expectation, and is not significant. There is no significant homology to human CDC2Hs⁷.

What might be the reason for this striking similarity between *E. coli* and yeast cell-cycle proteins? The functions of these proteins (other than the kinase activities

observed *in vitro* for CDC28 and CDC2^{8,9}), or their modes of action at the molecular level, are not known, but their similarity in this 60 amino-acid region suggests a common property. There is no direct evidence that FtsA shares the kinase function of the yeast proteins as the similarity found does not extend to the flanking regions in CDC28 and CDC2 that are identical to consensus sequences for ATP binding and phosphorylation sites. However, a nucleotide-binding motif¹⁰ (G---GK, residues 381–387) is found in FtsA distal to the region of similarity and a possible phosphorylation site (the dipeptide YT, residues 221–222) is proximal to the region. It is possible that this represents a reversal of the domain order between FtsA and the yeast protein kinases. Such domain order reversals have been postulated to occur in a ribosomal protein¹¹.

We may be dealing with a case of convergent evolution, as *E. coli*, *S. cerevisiae* and *S. pombe* are genetically highly diverse. Another possibility, is that this 60 amino-acid region arose in a common ancestor before the emergence of the eukaryotes. In that case changes in the domain's amino-acid sequence throughout the course of evolution will have been severely constrained by a constant functional requirement.

ARTHUR C. ROBINSON

Department of Medicine, University of Edinburgh,
The Royal Infirmary,
Edinburgh EH3 9YW, Scotland, UK

JOHN F. COLLINS

WILLIAM D. DONACHIE

Department of Molecular Biology,
University of Edinburgh,
Edinburgh EH9 3JR, Scotland, UK

- Donachie, W.D. & Robinson, A.C. in *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology (ed. Neidhart, F.C.) 1578–1593 (American Society for Microbiology, Washington, DC, 1987).
- Robinson, A.C. *et al.* *J. Bact.* **160**, 546–555 (1984).
- Smith, T.F. & Waterman, M.S. *J. molec. Biol.* **147**, 195–197 (1981).
- Lyall, A., Hill, C., Collins, J.F. & Coulson, A.F.W. in *Parallel Computing '85* (eds Feilmeier, M., Joubert, G. & Schendel, U.) 235–240 (North Holland, Amsterdam, 1986).
- Lörincz, A.T. & Reed, S.I. *Nature* **307**, 183–185 (1984).
- Hindley, J. & Phear, G.A. *Gene* **31**, 129–134 (1984).
- Lee, M.G. & Nurse, P. *Nature* **327**, 31–35 (1987).
- Reed, S.I., Hadwiger, J.A. & Lörincz, A.T. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4055–4059 (1985).
- Simanis, V. & Nurse, P. *Cell* **45**, 261–268 (1986).
- Walker, J.E., Sarste, M., Runswick, M.J. & Gay, N.J. *EMBO J.* **1**, 945–951 (1982).
- Iljas, A. & Thirup, S. *Chemica Scripta* **26B**, 109–119 (1986).

FtsA	L N L V N E E I L Q L Q E K L R - Q Q G V - R H H - L A A G
CDC28	L Y L V F E - F L D L D L K - R Y M E G I P K D Q P L G A D
CDC2	L Y L V F E - F L D M D L K - K Y M D R T S E T G A T S L D
FtsA	- I V L T G G A R Q I - E G L A A C - A Q R V F H T Q V R
CDC28	- I V - K K F M M Q L C K G I A Y C H S H R I L H R D L K
CDC2	P R L V - Q K F T Y Q L V N G V N F C H S R R I I H R D L K

Comparison of amino-acid sequences from segments of the FtsA protein of *E. coli*² (residues 305–357), the CDC28 protein of *S. cerevisiae*⁵ (residues 84–138) and the CDC2 protein of *S. pombe*⁶ (residues 80–136). The sequences are aligned for maximum homology¹. Amino-acids which are identical in FtsA and one or both of CDC28 and CDC2 are shown in solid-lined boxes, with conservative amino-acid changes in dashed-lined boxes. Related amino acids are grouped as follows: P, A, G, S, T — neutral, weakly hydrophobic; Q, N, E, D — hydrophilic acid amine; H, K, R — hydrophilic, basic; L, I, V, M — hydrophobic; F, Y, W — hydrophobic, aromatic.

Fertilization events

SIR—Dale¹ misrepresents the state of knowledge on the mechanism of fertilization. The existence of a rapid electrical block to polyspermy is solidly supported by data from sea urchins, starfish, the marine worm *Urechis* and frogs, where it has been shown that a rapid positive-going shift in the egg membrane potential occurs in response to insemination, that positive membrane potentials during insemination inhibit sperm penetration, and that negative potentials promote polyspermy^{2,3}. The recent experiments^{4–7} cited by Dale do not cast doubt on this hypothesis.

In fact, Shen and Steinhardt⁴ confirmed that positive potential inhibits sperm entry and negative potentials promote it: when periods of positive potential were interrupted by transient (10–80 ms) shifts to negative potentials, the more negative the potential, the more probable was fertilization. Although 2–3 dozen sperm were bound to an egg at the time the negative voltage windows were applied, only 43 out of 145 eggs were fertilized and most of these were monospermic. This does not contradict the existence of an electrical polyspermy block, because the more negative the voltage and the longer the duration of the voltage window, the more eggs were fertilized, as would be predicted.

Furthermore, the incidence of polyspermy during the negative voltage windows was not less than expected. Eighteen of the 43 eggs that were fertilized during the voltage windows were scored for polyspermy: one was dispermic and the rest monospermic. When the data are analysed by the Poisson distribution an expected frequency of 3 per cent is calculated for dispermic eggs close to the observed frequency of about 2 per cent.

Dale himself states of his data⁵ "The present report does not directly address the question of a fast block". In the other experiments^{6,7}, he cites sperm concentrations were deliberately kept low to avoid the polyspermy that would otherwise occur in eggs voltage-clamped to -20 mV. Thus there are no recent data that cast doubt on the occurrence of an electrical polyspermy block. To avoid confusion it should be pointed out that while some species possess electrical polyspermy blocks, others do not. Exceptions include hamster, the fish *Oryzias*, and salamanders (see ref. 2).

The other question raised by Dale was how sperm initiate egg activation. He suggested that if fusion and the first event in egg activation are simultaneous then it is more likely that the sperm injects materials into the egg through the cytoplasmic bridge formed at sperm-egg fusion than that there is the type of interaction in which sperm act as a ligand and the egg as a receptor. This is not logical. If fusion and the initiation of activation occur at the

same time, the most that can be concluded is that fusion could cause activation, but it is no less likely that there is another cause. In fact, we have shown in *Urechis* that protein isolated from sperm acrosomal granules causes activation of eggs, including the electrical response^{8,9}. Thus activation does not require sperm-egg fusion.

MEREDITH GOULD
JOSÉ LUIS STEPHANO

Escuela Superior de Ciencias,
Universidad Autónoma de Baja
California,
AP 1880 Ensenada, BCN Mexico

1. Dale, B. *Nature* **325**, 762-763 (1987).
2. Jaffe, L.A. & Gould, M. in *Biology of Fertilization* Vol. 3 (eds Metz, C.B. & Montroy, A.) 223-250 (Academic, New York, 1985).
3. Kline, D. et al. *J. exp. Zool.* **236**, 45 (1985).
4. Shen, S. & Steinhardt, R.A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1436 (1984).
5. Dale, B. *J. exp. Biol.* **118**, 85 (1985).
6. Longo, F. et al. *Dev. Biol.* **118**, 148 (1986).
7. Hinkley, R.E. et al. *Dev. Biol.* **118**, 148 (1986).
8. Gould, M. et al. *Dev. Biol.* **117**, 306 (1986).
9. Gould, M. & Stephano, J.L. *Science* **235**, 1654 (1987).

AIDS predictions

SIR—In criticizing our work, May and Anderson¹ not only refer to its presentation at the WHO meeting on the containment of AIDS (acquired immune deficiency syndrome) in Geneva when they mean the WHO meeting in Graz² but they also mistakenly characterize our work^{2,4} as an exercise in curve fitting. There is no joint work by us based on statistical analyses to fit polynomial or exponential curves to existing data on the incidence of AIDS. On the contrary, like May and Anderson¹, we have used the approximation of exponential growth for the number of human immunodeficiency virus (HIV) carriers to describe the initial stage of the epidemic. Taking into account the incubation time distribution we have studied the resulting incidence of AIDS. We have shown that the long and variable incubation times lead to transient phenomena characterized by an increase of the doubling times in the observable AIDS epidemic, of the incubation times, of the ratio of AIDS cases to HIV carriers, and so forth. As a consequence, the incidence of AIDS cases in the initial stage of the epidemic is nonexponential, notwithstanding the assumed exponential spread of the virus. In all our work we have stressed that the early exponential phase of growth of HIV carriers passes because of the depletion of the susceptible population and other inhibitory factors. We have looked at this depletion to provide upper limits for the duration of the exponential phase in various countries.

May and Anderson have a point in that extrapolation of trends gained by curve-fitting is unsafe, specially if done uncritically over an extended period. One of us (M.G.K.) made precisely this point in Anderson's presence at the Bilthoven meeting, December 1986. There are many

statistical analyses of AIDS based on curve fitting procedures with extrapolation over several years, some of them made at influential institutions (at the US Centers for Disease Control for example). Why, then, do May and Anderson aim their critical remarks only at our joint work, which has nothing to do with curve-fitting, and a very cautious projection for the United Kingdom⁵ made at a time when other prognostic analyses were scarce?

JOSE J. GONZALEZ
AID, N-4890 Grimstad, Norway
MICHAEL G. KOCH
V&C, S-546 00 Karlsborg, Sweden

1. May, R.M. & Anderson, R.M. *Nature* **326**, 137-142 (1987).
2. Gonzalez, J.J. & Koch, M.G. *AIDS Forsch* **11**, 621-630 (1986).
3. Gonzalez, J.J. & Koch, M.G. in *Proc. 1st Int. Meeting of AVIS on AIDS* (Solei, Milano, 1986).
4. Gonzalez, J.J. & Koch, M.G. *Am. J. Epid.* (in the press).
5. McEvoy, M. & Tillett, H.E. *Lancet* **ii**, 541-542 (1985).

Screwworm eradication and climate

SIR—I am surprised that Krafur and his newly acquired transatlantic colleagues¹ can dismiss my analysis² of the screwworm data so lightly. In respect of my equations (3) and (4), predicting seasonal numbers of screwworm cases in Texas from winter temperatures and summer temperatures, the regression coefficients for temperature are significant at the $P < 0.01$ and $P < 0.02$ levels, respectively, and crucially affect the model's predictability (see Fig. 4)². How can they dismiss my conclusions as "an artefact of pooling heterogeneous data", especially when they agree that "exceptional weather can have dramatic effects on screwworm incidence"?

In fact, if my model is applied to their new data for the seven separate climatological divisions in Texas, assuming that autumn cases (A) represent on average half the preceding summer's cases (the actual proportion does not matter), then the number of winter cases (W) in any particular region can be predicted from winter temperature by equation (1) in

Table 1. The parameters are clearly significant (see box), and, in fact, winter temperature looms larger in importance than autumn cases in the prediction ($R^2 = 0.49$, and 0.44 , respectively). In the winter to summer model, equation (2), summer temperature is just short of being significant ($P > 0.05$), but that is not surprising in such a small data set involving highly mobile fly populations.

Moreover, since publishing my analysis I have stumbled across independent evidence in support of the idea that change in climate rather than the release of sterile males might be responsible for screwworm eradication.

In their study of case incidence in various counties of South Texas in 1975-76, Krafur and Garcia³ inadvertently provide a fix on the overwintering temperature threshold for screwworm. In terms of reported cases, overwintering was possible only in the western counties of Webb, Zapata, J. Hogg, Star and Hildago where winter temperatures are about 2.7°C warmer than the average for South Texas as a whole (see graphs and p.691 in ref. 3). The average winter temperature for South Texas in 1975-76 was in fact 14.4°C (see Fig. 2 of ref. 3) so that an overwintering threshold of about 17.0°C is indicated.

This implies that the species would have been unable to overwinter even in southernmost Florida in 1957-58 when the mean winter temperature at Miami, for example, was only 16.7°C, the coldest on record in nearly 100 years. Similarly, in more recent times, the species could not have overwintered at, for example, Brownsville in southern Texas in 1977-78 or 1978-79, when the outbreaks collapsed and when the winters were the second and third coldest on record, nor in the winter of 1976-77 or any of the winters from 1982 to 1985, which were also very cold. More significantly, in relation to the current campaign in Mexico, the species would have experienced difficulty

	estimate	±	s.e.m.	t	P
a (intercept)	-22.94		3.66	6.26	<0.01
b (autumn cases)	1.49		0.34	4.40	<0.01
c (winter temperature)	0.28		0.05	5.40	<0.01

Table 1 Total reported cases of screwworm and average temperature for seven regions in Texas over 22 years¹

Region	Screwworm cases			Temperature (°F)	
	Autumn* (A _{n-1})	Winter (W _n)	Summer (S _n)	Winter (wt _n)	Summer (st _n)
Low Plains	2,010	2	4,020	43.2	81.4
Trans Pecos	4,635	11	9,270	46.8	80.1
East Texas	1,434	0	2,868	47.6	80.9
Edwards Plateau	11,475	137	22,951	47.8	81.3
South Central	8,631	103	17,263	53.4	82.9
Southern	9,633	830	19,265	55.4	84.6
Lower Valley	1,670	194	3,340	59.7	83.9

* Autumn cases assumed to be 1/2 summer cases.

$\ln(W_n + 1) = -22.94 + 1.49 \ln(A_{n-1} + 1) + 0.28 \ln wt_n$, $R^2 = 0.93$, $P < 0.001$. (1)

$\ln(S_n + 1) = 41.16 + 0.45 \ln(W_n + 1) - 0.41 \ln st_n$, $R^2 = 0.68$, $P < 0.05$, st_n not significant, $P > 0.05$. (2)

in overwintering as far south as Tampico in 1977–78, 1978–79 and 1983–84.

Thus, eradication from Florida in 1957–58 and the virtual disappearance of screwworm from the southwestern United States since about 1979, must in part have been due to unfavourable climate. The coincidence between 'success' in the two major campaigns and the occurrence of record cold winters is inescapable and unlikely to be due merely to chance. New outbreaks could occur when climatic conditions again favour the pest.

J.L. READSHAW

CSIRO, Division of Entomology,
GPO Box 1700,
Canberra, ACT 2601, Australia

1. Krafur, E.S., Townson, H., Davidson, G. & Curtis, C.F. *Nature* **323**, 495–496 (1986).
2. Readshaw, J.L. *Nature* **320**, 407–410 (1986).
3. Krafur, E.S. & Garcia, L.J. *med. Ent.* **14**, 687–697 (1978).

KRAFSUR ET AL. REPLY—In our earlier letter¹ we showed that the effects of winter temperature on screwworm abundance were statistically insignificant over 22 years in each of seven climatic zones of Texas. We do not think this can be fairly described² as a light dismissal of Readshaw's views³ that screwworm eradication by sterile males is a "grand delusion" and that the reduction in cases can be explained by recent cold winters.

Bartlett's test applied to the residual variances in the analysis of the data from the climatic zones shows that the pooled regressions are heterogeneous ($\chi^2 = 40.3$, 5 d.f., $P < 10^{-6}$). If, nevertheless, one chooses to pool the number of cases for the whole of Texas and take state-wide temperature means for each winter, there is, as Readshaw says, a statistically significant association of screwworm cases with winter temperature; we find that $F = 8.62$, $P = 0.009$, and that this association accounts for 13.5% of the variance. But we doubt whether it is wise to take a fairly weak association of this kind for such a large and biogeographically heterogeneous state and promote it to a causal predictive indicator. If the association were meaningful, one would expect it to be sharpened by looking for it zone by zone, but instead it completely disappears.

We are not convinced by Readshaw's use² of a significant screwworm/winter temperature association between zones averaged over all years to "explain" variation from year to year — this kind of manipulation of statistical association is notoriously unreliable. We also believe his "overwintering temperature threshold" of 17°C to be fallacious. He derived this idea from a paragraph⁴ containing data about the number of winter cases in different counties of southern Texas. But data in the same paragraph show that screwworm ovaries continue to mature even if temperatures drop to a mean of 15°C, which is scarcely consistent with the idea that at or below 17°C a population is

wiped out.

In contrast with Readshaw's belief in the irrelevance of sterile male releases, the same paper⁴ gives evidence that the manner of distributing sterile flies is critical to their achieving matings with wild females. Appropriate modifications to the sterile fly distribution techniques appear to have been decisive in reversing the fortunes of the eradication programme in 1977 in south and central Texas and in Tamaulipas, Mexico⁵, in contrast to west Texas, New Mexico and Arizona, where these modifications were not made at that time.

Readshaw's model (see Fig. 4 of ref. 3) predicts that there should be plenty of screwworm cases in Texas, but in fact none has been reported since 1982 (ref. 6). This follows the massive sterile male release campaign throughout Mexico over the past 10 years. This campaign seems to us much more likely to be the major cause of the eradication than do a few cold winters. By 1985 screwworms had been confined to east of the 93rd meridian in southern Mexico⁶. In March 1987 the Mexico–American Commission for the Eradication of the Screwworm reported that the screwworm was now restricted to the international frontier of Mexico with Guatemala and Belize, and that no screwworm cases had been found in the Yucatan peninsula since 21 December 1986 (S.A. Cajero and J.E. Novy, personal communication). We have examined winter temperature data for a number of locations in Mexico for 1936–86 and can see no trends that correlate with the recent screwworm eradication. In the lowland sites for which we have data, winter temperatures remained well above Readshaw's supposed 17°C threshold throughout the period of the eradication.

E.S. KRAFSUR

Department of Entomology,
Iowa State University,
Ames, Iowa 50011, USA

W.C. BLACK

Department of Biology,
Notre Dame University,
Notre Dame, Indiana 46556, USA

M.F. BRYAN

Statistical Laboratory,
Iowa State University,
Ames, Iowa 50011, USA

C.F. CURTIS
G. DAVIDSON

London School of Hygiene
and Tropical Medicine,
Keppel Street, London WC1E 7HT, UK

H. TOWNSON

Liverpool School of Tropical Medicine,
Pembroke Place, Liverpool L3 5QA, UK

1. Krafur, E.S., Townson, H., Davidson, G. & Curtis, C.F. *Nature* **323**, 495–496 (1986).
2. Readshaw, J.L. *Nature* **267**, 768 (1987).
3. Readshaw, J.L. *Nature* **320**, 407–410 (1986).
4. Krafur, E.S. & Garcia, L.J. *med. Ent.* **14**, 687–697 (1978).
5. Hofman, H.C. *Misc. Publ. Ent. Soc. Am.* **62**, 41–48 (1986).
6. Krafur, E.S., Whitten, C.F. & Novy, J.E. *Parasitol. Today* **3**, 131–137 (1987).

The multiple star system Sk –69 202

SIR—The star Sk –69 202, which is now firmly established¹ as the progenitor of SN 1987A, has two close companions. These companions^{2,3} have $V = 15.3$, separation, $d = 2.65$ arc s (star 2) and $V = 15.7$, $d = 1.4$ arc s (star 3). The joint *a priori* probability that Sk –69 202 should have two such relatively close companions, is $\sim 1 \times 10^{-5}$. This shows that Sk –69 202 was almost certainly the brightest member of a physical triple system. If star 2 is an unevolved main-sequence star it has a mass of $\approx 10 M_{\odot}$. As Sk –69 202 was brighter and more highly evolved than star 2, its main-sequence mass must have been $> 10 M_{\odot}$.

White and Malin² have derived star densities from various photographic surveys of the region within ~ 30 arc min of SN 1987A. From these data they calculate the probability that stars of various magnitudes will occur near an arbitrary point in this field. From their data it is found that the *a priori* probability of finding a star as bright as star 2 within 2.65 arc s of Sk –69 202 is $10^{-2.3}$ and the joint probability of finding stars 2 and 3 so close to Sk –69 202 is therefore 1×10^{-5} . In other words it is virtually certain that Sk –69 202 was the brightest component of a physical triple system. For an assumed distance of 50 kpc the projected separation of Sk –69 202 and star 2 is 0.6 pc. Because members of physical groups of stars are approximately coeval star 2 must be less evolved than Sk –69 202 and may therefore be used to set a lower limit on the mass of the supernova progenitor. A similar argument has recently been used by van den Bergh and Dufour⁴ to derive a lower limit for the mass of the progenitor of the supernova remnant N63A in the Large Magellanic Cloud, which is located in the association NGC2030.

S. VAN DEN BERGH

Dominion Astrophysical Observatory,
5071 W. Saanich Road,
Victoria, British Columbia,
V8X 4M6, Canada

1. West, R.M., Lauberts, A., Jørgensen, H.E. & Schuster, H.E. *Astr. Astrophys.* **177**, L1 (1987).
2. White, G.L. & Malin, D.F. *Nature* **327**, 36 (1987).
3. Chu, Y.H. *IAU Circ. No.* 4322 (1987).
4. van den Bergh, S. & Dufour, R.J. *Pub. astr. Soc. Pacific* **92**, 32 (1980).

Erratum

A different perspective on coral population genetics

R. Fitzhardinge

Nature **323**, 109 (1986).

Two references were incorrectly cited. They are correct as printed here.

1. Hodgson, G. *Mar. Ecol. Prog. Ser.* **26**, 61–71 (1985).
7. Maragos, J.E. thesis, Univ. Hawaii (1972).

Piecatcher!

John Rivers

The Politics of Food. By Geoffrey Cannon. *Century Hutchinson: 1987. Pp. 565. £14.95.*

WHEN Britain lost an Empire, it gained a national paranoia. This phenomenon can be glimpsed in the popularity of exposés of almost unbelievable secret plots and evil doing. But it is also evident in the fact that such accounts often cut close to the bone, for the activities of the industrial and political octopus that is the British Establishment leave much to be desired, and can only be explained by the wish of that Establishment to defend above all its right to steer the ship of state with as little interference from the people as it can get away with. In his exposé of the connections between the food industry, scientists and politicians, Geoffrey Cannon illuminates both sides of this paranoid equation.

Cannon is a seasoned campaigner for improvements to our diet, which he believes is at the root of much ill-health. He came into this field nearly ten years ago as the journalist who published the selective leaks from the NACNE committee, which forced an end to the deadlock between its members and allowed one particular view to triumph in the committee report. At the time Cannon made it clear in his articles, as he does in his recounting of the episode in this book, that, in NACNE, food industry and government representatives were conspiring to suppress the publication of a scientific consensus. I have no doubt that there was a conspiracy, but there is more to the story than Cannon tells us.

NACNE was the British analogue to the McGovern Committee in the United States, though with typically British irresolution it was constituted as a non-governmental, semi-official body. The letters stand for the National Advisory Committee on Nutrition Education, with the National only added when it was realized what the acronym otherwise meant. Theoretically it existed to provide guidelines on nutrition education to the equally semi-official Health Educational Council, since abolished by the government. In practice, largely thanks to the publicity generated by Mr Cannon and others, NACNE was the body whose stringent proposals revolutionized the British idea of healthy eating. Its endorsement of low-fat, high-fibre, low-salt diets gained public attention where all previous attempts had failed.

The reason for that success, I suspect, lay not in the superior nutritional wisdom

or arguments of the committee, nor in an improvement in the tenuous scientific evidence that previous committees had found insufficient, but in our national paranoia. The Great British Public believe anything if they are told that the food industry and government are conspiring to suppress it. Thus the NACNE view of nutrition and disease triumphed, with public and health professionals alike, largely because of the well-orchestrated and continuing press campaign which now makes doubters (such as myself) appear to

"Readers outside Britain will be astounded that the committees dealing with matters of food and health are covered by the notorious Official Secrets Act. But then in Britain the time of day will soon be an official secret."

be tainted because of their cooperation with the food industry.

In its recruitment of the press to its cause, the health lobby has only adopted the public-relations approach to the issue of food and health so long used in Britain, and other Western nations, by such bodies as the Butter Information Council and the Sugar Bureau. Personally, I don't like it, and find public relations no more palatable when practised by the knights of the round table than by anyone else. But one may have to accept such a system, for what is going on is not a scientific battle for truth, but a political battle in which industry and civil servants are broadly defending the *status quo* against the dietary reformers, each side trying to make its own beliefs the most politically palatable.

Mr Cannon, though, is not insincere in what he says. Indeed as this book testifies at great length and in a semi-scholarly manner, he remains unshaken in his belief that the truth is clear and a conspiracy exists to suppress it. But in trying to detail why there is a conspiracy, or how it works, he is less than convincing. It seems quite proper to me that politicians should oppose unnecessary change. Of course the food industry exists to make a profit, and to avoid the implication that its products cause harm will argue against tentative evidence. Some companies may be indifferent to proven ill-health that their product may cause, but a failure to believe every accusation is not evidence of such

callousness. Junk food is not thalidomide.

In Geoffrey Cannon's schema, the evidence of harm seems to be widespread and all but irrefutable, so the callousness is self-evident in every ministerial equivocation. Why then, he asks, does not reform occur? The answer it seems must be malice. Eventually, he is led to suggest that the food and pharmaceutical industries may collogue to keep salt consumption high. This is a black and white view of things, insensitive to the greys of human behaviour or the muddiness of political choices. Yet Cannon is right to look with disdain on the posturing of many nutritional scientists in Britain, to suspect that their intimate association with Whitehall and the food industry leads to complacency. He is right to be concerned about the implications of cosy cabals of chums working in secret to produce an expert committee report, often based on unreferenced working papers. The whole structure is overdue for reform.

The problem with expert committees is a general one affecting the United Nations and many national governments. The British may simply have it worse than others. Readers outside Britain will no doubt be astounded to learn that the considerations of the committees dealing with matters of food and health are covered by the notorious Official Secrets Act. But then in Britain the time of day will soon be an official secret. The official excuse — that such committees may deal with trade information which industry will only disclose if kept confidential — is a feeble one, but Cannon's implication that the secrecy implies shady goings-on does not follow. Although the Whitehall secrecy blanket may make for flabby administration, it doesn't necessarily imply widespread corruption. Certainly, however, it does achieve an increase in paranoia, and provides, in the selective leak, a powerful weapon for railroading committees.

Cannon is very concerned with venality, and lists at great length the food industry contacts of many scientists who sit on expert committees, as well as of the politicians they advise. Though he carefully denies that he is making any accusation of less than totally honourable behaviour, he actually manages to imply at crucial points that this contact reduces the objectivity of scientists and somehow disqualifies them from advising. It is a common viewpoint amongst food reformers, but a two-edged one. Because there is considerable overlap between the experts who advise government and those who approve government grants, one could argue that commercial funding demonstrates the independence of the investigator from government influence. Cannon does not pursue this line, presumably because he believes that the food industry and government interests are one and the same.

Moreover I wonder if he might be secretly glad that these commercial contacts exist, for the *ad hominem* technique is a powerful one: if there is no evidence against a scientist's research, look at who paid for it.

It is this question of the validity of research that is the missing heart of *The Politics of Food*. The food and health debate is complicated because it is about probabilities not truths, but discussion of the issue of how such scientific evidence should be evaluated is squeezed into the end of the book. Cannon's treatment of this problem, though valuable, is sketchy. His proposal that expert committees should function in a quasi-judicial capacity, because their "job . . . is to produce a verdict based either on the standard of proof required in a civil court (balance of probability) or else in a criminal court (beyond reasonable doubt)", is sensible but too skeletal. How is the presumption of innocence to work: is a theory, or a set of experiments, to be assumed true until falsified, or to be assumed false until the burden of confirmation overwhelms administrative conservatism? Given that scientific evidence is to be incorporated into policy with far less objective political judgements, perhaps the very criteria for proof should be reconsidered.

One cannot help but suspect that in this respect Cannon lacks a feel for what scientists actually do. What he seems not to appreciate is that all scientific evidence is not equal, that even the unprejudiced reviewer does not give equal weight to all scientific reports, but rejects some as isolated, unconfirmed or even implausible.

This lack of scientific expertise is evident in the discussion of the technical as well as the political aspects of food policy, in what is by far the weakest section of the book. Here Cannon is sometimes in error — as when he states that the azo-dye amaranth is inorganic — or he misinterprets as in his idiosyncratic, and invalid, use of recommended daily allowances. And he is an ardent consumer of the unconsidered fragments of the scientific literature which he uses, for example, to argue that anorexia nervosa is in fact a manifestation of widespread zinc deficiency. This armchair science, which is largely presented in the manner of a pulp detective story, detracts from his more valuable socio-political critique.

Faults such as this, together with the lapses into an irritating, journalistic prose style, mean that this potentially valuable book has to be counted as a wasted opportunity. Sadly, it is likely to receive much less serious attention than it, or its author, deserve. □

John Rivers is a Senior Lecturer in the Department of Human Nutrition, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK.

Island progress

Ernst Mayr

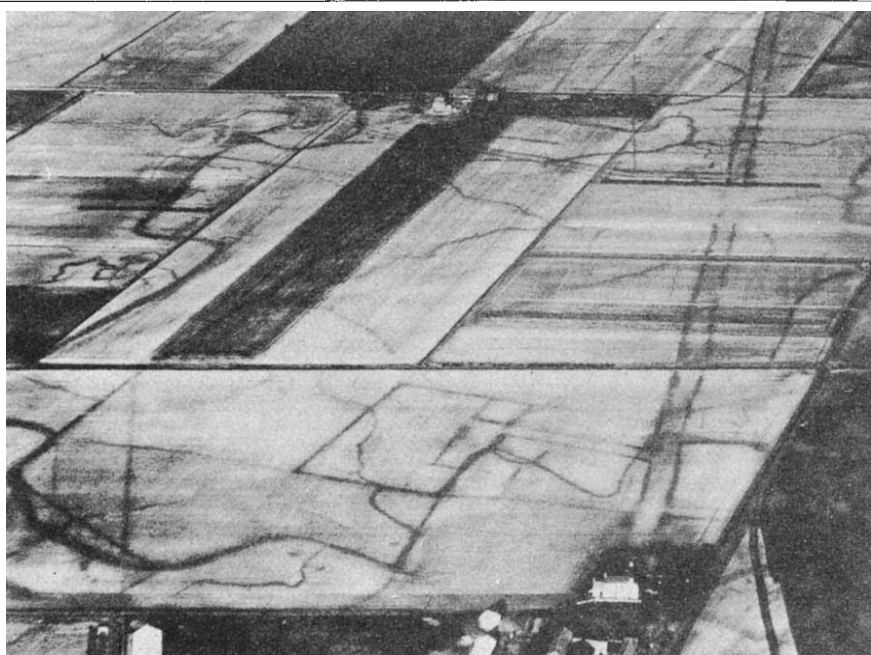
The Malay Archipelago. By Alfred Russel Wallace. Introduction by John Bastin. Oxford University Press: 1987. Pp. 638. £27.50, \$35.

THAT Alfred Russel Wallace was the co-discoverer with Charles Darwin of the principle of natural selection is now known to every educated person. That he was also one of the greatest travelling naturalists the world has ever seen, deserves, however, to be known far more widely. To reprint *The Malay Archipelago*, his famous travel book, was therefore a happy idea. For eight years (1854–62) Wallace criss-crossed the archipelago between Borneo and Java in the west and New Guinea and the Aru Islands in the east. No modern explorer can conceive of the hardships and dangers of travelling in small native boats among usually unhelpful, if not hostile, people. Having myself once sailed for nine months in these waters in a motorless boat, I know how it feels when the wind suddenly dies down and the current carries one's boat rapidly toward the breakers on the rim of a coral reef. It is a miracle that Wallace did not perish on one of these occasions. After reading *The Malay Archipelago* Darwin was moved to write him: "Of all the impressions which I have received from your book, the strongest is that your

perseverance in the cause of science was heroic".

With an unquenchable enthusiasm and dogged determination Wallace continued to explore island after island, not only discovering thousands of new species, but revealing for the first time the intricate intermingling of the Asiatic and Australian faunas in this island region, appropriately now called Wallacea by the biogeographers. It was Wallace who discovered the sharp break between the rich faunas of the continental islands (for example Borneo, Sumatra) on the continental shelf and the impoverished faunas on islands such as Sulawesi, the Lesser Sunda Islands and the Moluccas, which had never had any continental connections. Being strictly isolated and receiving only occasional immigrants by across-the-water colonization, these islands are extraordinarily rich in endemics. Because he was the first naturalist to visit many of the islands, or at least the first to make systematic collections, Wallace was able to make an astonishing number of new discoveries, including 212 new species of birds, 900 new kinds of beetles and hundreds of new butterflies, ants, other insects and snails.

But Wallace, the naturalist, was of course far more than a collector. He constantly searched for the meaning of what he observed, and wrote two of his most important scientific papers while in the archipelago. In his 'Sarawak paper', written and published in 1855, he advanced abundant evidence in favour of



Ghostly shadows — two eras of Fenland farming are seen superimposed, to the south-west of Spalding, Lincolnshire. The modern landscape is of arable farming and drainage ditches, the Roman pattern emerges in the form of soil marks visible from the air. Above and to the left of the modern farm is a rectangular Roman house plot, connected to the through-road by a short, straight access drive. The illustration is taken from *History from the Air* by Richard Muir, originally published in 1983 and just issued in paperback by Mermaid Books, with photographs from the University of Cambridge collection of air photographs. Price is £10.95.



Wallace — a truly lovable person.

geographic speciation, thus decisively undermining the dogma of the constancy of species (four years before *The Origin of Species* was published). And early in 1858 in the Moluccas he drafted the famous essay in which he developed a theory of evolution based on natural selection, subsequently published by the Linnean Society together with excerpts from Darwin's manuscripts.

No reference to these writings is made by Wallace in *The Malay Archipelago*, yet his broad interests are as apparent on every page as in Darwin's *Journal of Researches* (Beagle). Wallace was particularly intrigued by the diversity of human populations. For every island he visited, he reports what racial type the natives were, what language they spoke, what character they had, how they lived and what crops they raised. He was deeply concerned over the adverse impact of civilization on their lifestyle and wonders sorrowfully what their future would be.

In these reflections Wallace's genuine goodness (there is no better term) is fully apparent. He was quite a wonderful person — always modest, never complaining even when everything went wrong, generous to a fault and full of ideals, something that also emerges in his later books. If there ever was a truly lovable person, it was Alfred Russel Wallace.

This new edition of *The Malay Archipelago* is greatly enriched by an excellent, extensively researched introduction by John Bastin. It tells us much about Wallace's life, the history of his writing the book, the nature and extent of Wallace's discoveries, the book's success (number of editions, and so on), the exact itinerary of his 14,000 miles of travelling and the recent Wallace literature. We are much indebted to Mr Bastin for providing such an informative introduction to his new edition of Wallace's immortal work. □

Ernst Mayr is Alexander Agassiz Professor Emeritus at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Molecules and man

David Weatherall

Molecular Biology of Homo Sapiens. Cold Spring Harbor Symposia on Quantitative Biology, Vol. LI. *Cold Spring Harbor Laboratory*: 1987. Two volumes, pp. 1,299. Hbk \$160; pbk \$80.

THE titles of the first fifty Cold Spring Harbor Symposia on Quantitative Biology reflect a period of change and development in the biological sciences quite as remarkable as the great years of physics that preceded it. This excellent series covers some of the early applications of chemistry to the study of cellular function, the development of modern protein chemistry and immunology, the discovery of the structure of DNA, the genetic code and the mechanisms of protein synthesis, and, most recently, the emergence of recombinant DNA technology and its application to every branch of biology. The only previous symposium devoted to human biology was held in 1964. Entitled "Human Genetics" it was confined to population studies, human proteins, and somatic cell genetics, and as I recall was a low-key affair with little sense of the exciting developments that were just round the corner.

The size and weight of the two books that contain the proceedings of the fifty-first symposium, devoted entirely to human molecular biology, emphasizes the remarkable progress that has been made in this field in recent years. The opening sections of the first volume are devoted to the human gene map. In a lively introduction Bodmer makes a strong case for Project 2000, the establishment of the complete sequence of the human genome by the end of the century. Subsequent chapters cover genetic diagnosis and the development of new methods for determining the molecular pathology of single gene disorders, molecular evolution, and the generation of growth factors and other regulatory proteins by recombinant DNA technology. The second volume encompasses receptors, cancer, and prospects for gene therapy, and ends with a useful summary of the meeting by Caskey and an appendix collated by McKusick that brings together current information about the human gene map.

The organizers must be congratulated on planning this large symposium in a way that provides such excellent coverage of this rapidly expanding field. Although much of the material has been published elsewhere, its appearance together in a series of well-written reviews emphasizes the extraordinary power of recombinant DNA technology for human biology and medicine. The scope of the field is immense, ranging from the determination

of the cause of diseases of completely unknown aetiology by reverse genetics, through the molecular mechanisms of sex determination and colour blindness, to the analysis of human evolution and the origin of races by the study of polymorphisms of chromosomal or mitochondrial DNA. The sections on cloning and determining the functions of growth factors are particularly well timed; just as these volumes went to press the first results of clinical trials of genetically engineered erythropoietin for the treatment of the anaemia of renal failure appeared in the medical press. The outcome, which must have allowed those with shares in the biotechnology industry to have a decent night's sleep at last, was a clear vindication of the practical value of recombinant DNA technology for medical practice. And another remarkable story that is well aired is how the revitalized field of cytogenetics has combined forces with molecular immunology and the cell biology of oncogenes to yield some tantalizing clues about the pathogenesis of malignant transformation.

Of course, in reviewing an enterprise of this scope it is always possible to find omissions or apparent imbalances. For example, a great deal of space is devoted to gene therapy, still very much at the experimental stage, and yet there is very little discussion of the practical aspects of DNA analysis for the early prenatal diagnosis of single gene disorders, surely one of the major clinical success stories of this field. And since one of the main applications of recombinant DNA technology for medical practice in the future may well be for sorting out the molecular pathology of common conditions such as vascular disease, diabetes, autoimmunity and the major psychoses, it is disappointing to see that, with the exception of heart disease, none of these polygenic conditions is discussed. But overall these superbly produced volumes provide by far the most complete account of the state of the art of human molecular biology and its long-term applications that is currently available.

The list of speakers and participants at this meeting is an interesting reflection of the way in which medical research is changing now that recombinant DNA technology has come to stay. No less than 28 of the chapters are from biotechnology companies. Nearly all the papers have long lists of authors with multiple institutional affiliations, at least some of which reflect highly productive collaborative ventures between university departments and industry. And while some of the work reflects individual innovation, the overall impression is of a field being swept along by large collaborative teams using similar technology; given its current impetus there is no reason why this approach should not enable it to achieve its next

major *tour de force*, the sequence of at least the interesting parts of the human genome by the year 2000. But it is beginning to look as if we shall have to wait a lot longer before we know how the whole thing works; perhaps project 3000 will tell us how a fertilized egg turns into a human being.

These thoughts are echoed in the foreword to these volumes, in which Watson predicts that the overwhelming information explosion in this field, rather than reflecting a short period of highly productive research in the human biological sciences, gives every indication of opening up new frontiers that will take hundreds if not thousands of years to explore thoroughly. It is slightly disturbing, therefore, that the symposium does not include any discussion of the broader social implications of the new genetics or any mention of the ethical questions that may arise as we come to dissect the human genome down to its last few molecular bricks. How, for example, will doctors cope with an increasingly reductive

approach to the medical sciences when, at the same time, they are being urged to concentrate more on the holistic aspects of patient care? And how far do we wish to go in the social engineering that will follow inevitably from our increasing knowledge of the human genome? Surely it must have been the same in the physical sciences at the beginning of the century; things were moving so fast and so much excitement was being generated that there was no time to step back and evaluate the long-term implications of what was happening. But if the biologists are not to create problems of even greater magnitude than those that stemmed from the vintage years of physics, it will be important for us to reflect on these questions; given the current pace of development of human molecular biology, as evidenced by these splendid books, this will not be easy. □

Sir David Weatherall is Nuffield Professor of Clinical Medicine and Honorary Director of the Medical Research Council Molecular Haematology Unit, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Hardy perennial

Clive Stace

Flora of the British Isles, 3rd Edn. By A.R. Clapham, T.G. Tutin and D.M. Moore. Cambridge University Press: 1987. Pp.688. £65, \$125.

Flora of the British Isles (or 'CTW' as it is almost invariably known) is the standard British and Irish flora, and the appearance of the third edition 35 and 25 years after the first and second editions, respectively, is a noteworthy event. The accounts prepared by E.F. Warburg, who died in 1966, have been taken over and updated by D.M. Moore of Reading University.

Although the book looks quite different at a glance, with its glossy front and new text format in two columns, this third edition is clearly only an updated version of its predecessors, and in some respects only partially updated at that. Hence all the long-familiar features of CTW are still to be found and need not be described again here. Its strongest points are the reasonably detailed descriptions and the relatively copious peripheral data included on such points as ecology, pollination, phenology, dispersal and chromosome numbers. Mention of a large number of commonly grown related garden plants is also an often welcome additional feature. This flora is therefore of value to a wide range of biologists — in taxonomic terminology it is a general-purpose flora rather than a specialist's handbook.

Since the second edition of CTW (published in 1962), all five volumes of *Flora Europaea* (1964–1980) have appeared,

and, as expected, many of the latter's taxonomic features are utilized in the latest edition of CTW. The preface states that national floras in Europe should adopt the taxonomy and nomenclature of *Flora Europaea*, "unless there seemed good grounds for doing otherwise". In fact there are quite frequent deviations (but the "good grounds" are not explained), for example *Alcea/Althaea*, *Chamaenerion/Epilobium*, *Filago* etc., *Centaurea nigra* agg., *Salicornia* species. The sequence of families is retained as in the earlier editions (except for the pteridophytes), but within the families the generic sequence has often been modified, frequently to agree with *Flora Europaea*.

To the specialist the greatest disappointments fall into three categories. There are many inconsistencies in the treatments, ranging from contradictions to stated policy (for example, names of counties and English names of plants used) to the very varied treatment of comparable groups (for example, three and ten columns, respectively, for the abundant *Taraxacum* and *Rubus* agamospecies, yet twenty-four for the much less common *Hieracium*; random use of terms 'agg.', 'group', and so on). Secondly, there are numerous examples of (mostly trivial) points that have not been adequately updated. Many names used have been shown to be incorrect according to the *Code* (such as *Thlaspi alpestre*, *Oenothera erythrosepala*, *Festuca tenuifolia*); distribution patterns are sometimes outdated (especially concerning aliens); results of some recent taxonomic research have not been incorporated. (*Rumex tenuifolius*, *Carduus acanthoides*, *Rhynchosinapis*,

among others); and the figures have not been updated since the 1952 edition (*Sorbus wilmottiana*, for example, described even in the 1962 edition, has not been added). Thirdly, the coverage of alien plants is very often unsatisfactory. Many species that no longer occur (for example, *Rosa sempervirens*) or never did (for example, *Phytolacca americana*) are given extensive coverage, while others now common, such as *Geranium x magnificum*, are not mentioned. Twenty lines are devoted to *Salix elaeagnos*, which is not naturalized in Britain, while *S. cordata*, well naturalized in Warwickshire, is not covered. Numerous similar examples exist.

Unfortunately, due to a serious clerical error, the family Zannichelliaceae has been completely omitted from the text and index, though it appears in the family keys and synopsis. After its point of omission (p.531) the families have been re-numbered, so that from that point their numbers do not correspond with those in the synopsis and keys. The common grass called *Festuca diffusa* in *Flora Europaea* and *F. rubra* ssp. *multiflora* in C.E. Hubbard's *Grasses* (Penguin, 3rd edn, 1984) is omitted altogether, probably because Britain was erroneously not included in its distribution in *Flora Europaea*.

Production of the 1952 edition of CTW was a magnificent achievement that has not been equalled since. The latest edition is likely to remain the standard British flora well into the next century, but it is a pity that it has not quite kept pace with our increase in knowledge. The price, however, at 26 times that of the original, is fully updated. □

Clive Stace is Professor of Plant Taxonomy in the University of Leicester, Leicester LE1 7RH, UK, and President of the Botanical Society of the British Isles.

New in paperback

- *Light and Photosynthesis in Aquatic Ecosystems* by John T. O. Kirk. Publisher is Cambridge University Press, price is £15, \$24.95. For review see *Nature* 309, 382 (1984).
- *An Introduction to Science Studies: The Philosophical and Social Aspects of Science and Technology* by John Ziman. Publisher is Cambridge University Press, price is £9.95, \$14.95. For review see *Nature* 314, 204 (1985).
- *Intermolecular Forces: Their Origin and Determination* by Geoffrey C. Maitland, Maurice Rigby, E. Brian Smith and William A. Wakeham. Publisher is Oxford University Press, price is £20, \$37.50. For review see *Nature* 294, 384 (1981).

New in Britain

- *The Griffin* by Arnold Kramish. Publisher is Macmillan, price is £12.95. For review see *Nature* 325, 203 (1987).
- *Prescription For Disaster: From the Glory of Apollo to the Betrayal of the Shuttle* by Joseph J. Trento. To be published on September 24 by Harrop, price £12.95. For review see *Nature* 327, 289 (1987).

Interaction and annihilation of antiprotons and nuclei

T. von Egidy

Physik-Department, Technische Universität München, D-8046 Garching, FRG

Experiments with slow antiprotons and nuclei reveal what happens when matter and antimatter come together and annihilate. Before the annihilation antiprotonic atoms are formed, and after the annihilation many fast particles are emitted and much lighter residual nuclei are left.

MORE than 50 years ago, the British physicist Paul Dirac introduced to physics the concept of antiparticles and antimatter. This turned out to be one of the most fascinating and fruitful ideas in modern physics. At first, this idea appeared very strange to Dirac's colleagues.

The idea of antiparticles arose from Dirac's trust in mathematics. When he established his relativistic quantum mechanical equation for the electron he wrote¹: "the wave equation has in addition to the wanted solutions for which the kinetic energy of the electron is positive, an equal number of unwanted solutions with negative kinetic energy for the electron, which appear to have no physical meaning". But one year later he suggested² that the 'unwanted solutions' are antielectrons which annihilate with electrons into two 0.5 MeV gamma rays, and proposed "presumably the protons will have their own negative-energy states, all of which normally are occupied, an unoccupied one appearing as an antiproton". A year later, in 1932, C. D. Anderson³ discovered the antielectron, called the positron, with the properties predicted by Dirac. For the production of antiprotons, accelerators with energies of several GeV were necessary, and sensitive and selective detection methods had to be developed. In fact the first antiprotons were only identified by Chamberlain, Segrè, Wiegand and Ypsilantis⁴ at Berkeley, California, in 1956. The first observation of antineutrons followed almost immediately by Cork, Lambertson, Piccioni and Wenzel⁵ at Berkeley. Even heavier antiparticles such as antideuterons⁶ and antihelium-3⁷ have been produced at Brookhaven, New York, and Serpukhov in the Soviet Union, respectively. It is now a well-established fact in particle physics that all particles with spin $\frac{1}{2}$ (fermions) including neutrinos have corresponding antiparticles with opposite electric charge. The antiparticles of the constituents of our normal matter, however, deserve special attention. Positrons are available to experimentalists through radioactive decay or pair production. In 1983 the purest and most intense source of antiprotons was commissioned at CERN in Geneva: the Low-Energy Antiproton Ring (LEAR).

A special feature of antiparticles are their negative particle numbers. Antiprotons and antineutrons have negative baryon numbers and positrons and antineutrinos have negative electron-lepton numbers. As these particle numbers are conserved, antiparticles can be created or annihilated only with particles of corresponding positive numbers.

Although physicists have not yet been able to put an antiproton and a positron together to construct an antihydrogen atom or to combine many antiprotons and antineutrons to create heavier antinuclei, there is no doubt that more complex antimatter could in principle exist, unless fundamental symmetries are to be violated. Therefore, it is expected that an antiworld would look essentially like ours, but with opposite electric charges. If all properties of particles and antiparticles are com-

pletely symmetric the question arises as to why our world is made of particles and not of antiparticles, and what happened after the Big Bang (the creation of our Universe), so that only matter survived.

Another problem is the question of whether matter and antimatter should gravitationally attract or repel one another. There are strong arguments⁸ for the attraction of matter and antimatter which basically depend on the observation that photons exhibit positive gravitation. Since a proton-antiproton pair can be converted into two photons and vice versa, the energy would not be conserved if the antiproton had negative gravitational mass. The gravitational forces on positrons and antiprotons have not yet been measured, as electric forces are many orders of magnitude stronger than gravitational ones. Presently a large American and Italian cooperation⁹ is preparing an experiment at LEAR/CERN to measure the gravitational mass of the antiproton, with a precision of better than 1%. Such precision is necessary to test speculations¹⁰ that the gravitational mass of antimatter is larger than that of matter.

Processes before and after annihilation

Turning our attention first to the 'simplest' antiparticles we observe that fast positrons are stopped in matter in a fashion similar to that of fast electrons. The subsequent annihilation of positrons with electrons results in two gamma rays of 511 keV, emitted in opposite directions.

The fate of antiprotons ($\bar{p}$) in matter is more complicated. Fast antiprotons can either annihilate in flight if they come close to a nucleus or can be stopped in matter just like protons (p) according to the Bethe-Bloch formula, (with many small energy transfers to electrons). If the antiprotons have energies of less than a keV, they are caught by an atom by means of their negative charge, entering an atomic orbital in the process. The first principal quantum numbers n of the antiprotons are $\sim \sqrt{m_p/m_e} = 43$ times larger than the n of the electrons they replace. Therefore, initial n -values larger than 100 are expected for heavier atoms. The antiprotons starting at these large orbits cascade down, first by emission of Auger electrons (electrons ejected by energy transfer) and later by emission of X rays. Finally, the orbits of the antiprotons are so close to the nucleus that they interact not only by the long-range electromagnetic forces but also by the short-range, strong nuclear forces. This takes place if $n \leq 4$ for oxygen or $n \leq 9$ for lead (more precisely if $l \leq 3$ or $l \leq 8$, respectively, where l is the angular momentum quantum number). The strong interaction between antiprotons and nuclei is not well understood yet, and is a field of active research using antiprotons. Information can be obtained either from antiprotonic atoms or, for higher energies, by the scattering of antiprotons by nuclei.

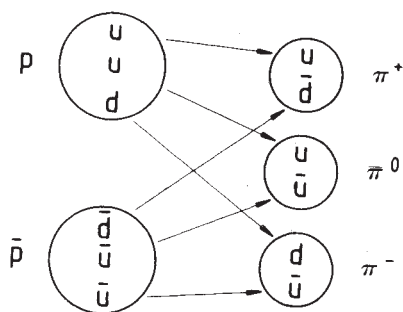


Fig. 1 Rearrangement of quarks is the most simple picture of the annihilation of a proton and antiproton into three pions.

The strong antiproton-nucleus interaction is dominated by absorption. Therefore, the slow antiprotons in antiprotonic atoms rarely penetrate the nucleus but annihilate with a proton or neutron at the nuclear surface. But the annihilation process inside the nucleus is very interesting, because it has been speculated that special nuclear matter states might be created by the well-localized, very high annihilation energy of 1.88 GeV. In this context, fireballs, large quark bags, special quark-gluon plasma, hotspots, hot nuclear gas, shock waves, density waves, the explosion of the nucleus, and the creation of hyperons have been discussed¹¹⁻¹⁶. The most simple way to understand the annihilation of an antiproton with a proton or neutron is the rearrangement of the three quarks in the proton or neutron with the three antiquarks in the antiproton. This yields three pions as shown in Fig. 1. In more than 95% of all low energy $\bar{p}p$ annihilations only various combinations of pions (π^+ , π^- , π^0) are produced, ranging from two to eight in number (on average five). The pions are produced either directly or they are decay products of mesonic resonances (η , ρ , ω). In a few per cent of the annihilations, kaons or hypernuclei are produced. Five pions have a rest mass of $5 \times 140 \text{ MeV}/c^2 = 700 \text{ MeV}/c^2$; therefore from the annihilation energy of 1,880 MeV $\sim (1,880 - 700)/5 \text{ MeV} = 236 \text{ MeV}$ of kinetic energy per pion is left over. From these pions some escape directly whereas others enter the nucleus and start an intranuclear cascade. During this cascade many processes take place: the fast pions are adsorbed or scattered in the nucleus; Δ -particles or additional pions are produced; protons, neutrons, deuterons, tritons or α -particles are knocked out; and finally the very hot nucleus evaporates neutrons and some protons. What is left is a nucleus, in most cases radioactive, which has lost many nucleons (protons and neutrons), sometimes >50 from heavy nuclei. Due to the preferred evaporation of neutrons, most of the residual nuclei are neutron deficient. Recently, hypernuclei have been reported¹⁷ after $\bar{p}$ annihilation in ^{238}U . These are nuclei with a Λ hyperon replacing a nucleon in the nucleus. But as yet, other 'special' nuclear matter states have not been identified after $\bar{p}$ -nuclei annihilations.

Future experiments using antiprotons will attack such problems. Where the theoretical predictions are sufficiently specific, exotic states may be detected by special correlations of emitted particles such as the multiplicity or angular correlation.

Production of antiprotons

The European research centre CERN at Geneva includes the most effective antiproton source in the world. The production of low-energy antiprotons with LEAR is the result of a fascinating combination of many accelerators and storage rings. The new developments for this system, including the high-energy 'branch', are so outstanding that van der Meer obtained the Nobel prize for physics in 1984 for his contributions to its design. Figure 2 displays the relevant facilities at CERN¹⁸. Protons are accelerated in the linac to 50 MeV, in the booster (PSB) to 800 MeV and in the proton synchrotron (PS) to 25,000 MeV

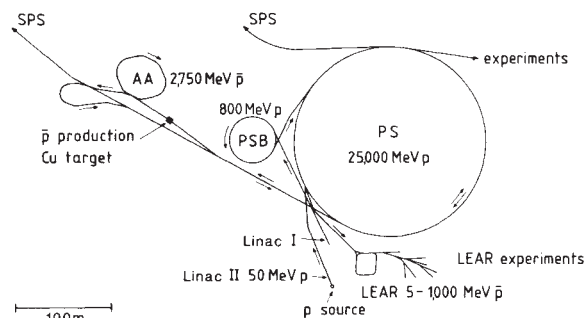


Fig. 2 The facilities at CERN, Geneva, used for the production, storage and handling of antiprotons (ref. 18).

(26 MeV/c). These protons, which have ~ 28 times their rest mass, are focused on to a copper target (11 cm long and 3 mm in diameter), where the antiprotons are produced. The antiprotons are collected in the antiproton accumulator (AA). About two million fast protons are needed to generate one antiproton stored with 2,750 MeV (3.57 GeV/c) in the AA (accumulation rate about $6 \times 10^9 \bar{p}$ per hour). These antiprotons are 'cooled' (the volume of phase space that they occupy, primarily the transverse momentum component, is reduced). About once per hour a stack of antiprotons is kicked out, transferred to the PS, where they are decelerated to 175 MeV (0.6 GeV/c) and sent to LEAR. LEAR stores, cools and decelerates (or eventually accelerates) antiprotons and delivers 5-1,000 MeV (0.1-1.7 GeV/c) antiprotons at a rate of $\sim 5 \times 10^5 \text{ s}^{-1}$ for experiments.

A new antiproton collector (ACOL) is presently under construction at CERN, which will have accumulation rates ten times greater than before. LEAR is also being improved. These new facilities should be available by the autumn of 1987.

Scattering of antiprotons by nuclei

A large magnetic spectrometer has been installed at LEAR to measure elastic and inelastic scattering of antiprotons by nuclei as a function of the scattering angle¹⁹. In one experiment, antiprotons with energies of 47 MeV and 180 MeV were scattered off ^{12}C , ^{16}O , ^{18}O , ^{40}Ca and ^{208}Pb targets, momentum-analysed by two dipole magnets and detected in multiwire proportional counters. The energy resolution was 1 MeV. The main purpose of these experiments was the determination of the $\bar{p}$ -nucleus optical potential. Figure 3 (ref. 20) shows the differential cross-section for elastic scattering of antiprotons by ^{12}C as a function of the centre of mass angle. Values for elastic proton scattering are added for comparison. A strong diffraction pattern is visible. Optical potentials with a strong absorptive imaginary part (61 MeV) and a weaker attractive real part (25 MeV) reproduce the experimental results. A recent careful analysis²¹ of the $\bar{p}$ scattering data showed that the potential is well-defined only at the surface of the nucleus due to the strong absorption of the antiprotons. For 47 MeV $\bar{p}$ the absorption takes place where the nuclear density is $<10\%$ of the central value and at those radii the real potential is $\sim 50\%$ of the imaginary one. Therefore, the inner details of the nucleus plays no role.

Previous ambiguities concerning the optical potential²² have now been solved. The shallow, attractive and deep absorptive potential also excludes antiprotons orbiting in the strong potential around the nucleus, an idea proposed recently^{23,24}.

Strong interaction in antiprotonic atoms

Shown in Fig. 4 is a scheme of energy levels in the antiprotonic atom Nd. The levels are characterized by the principal quantum number n and the angular momentum quantum number l . The antiprotons jump to lower levels and predominantly emit Auger electrons for higher n -values ($n \geq 15$ for Nd) and X rays for lower n -values. If the angular momentum l is small ($l \leq 8$ for Nd), the antiprotons approach the nucleus within a few fm, the

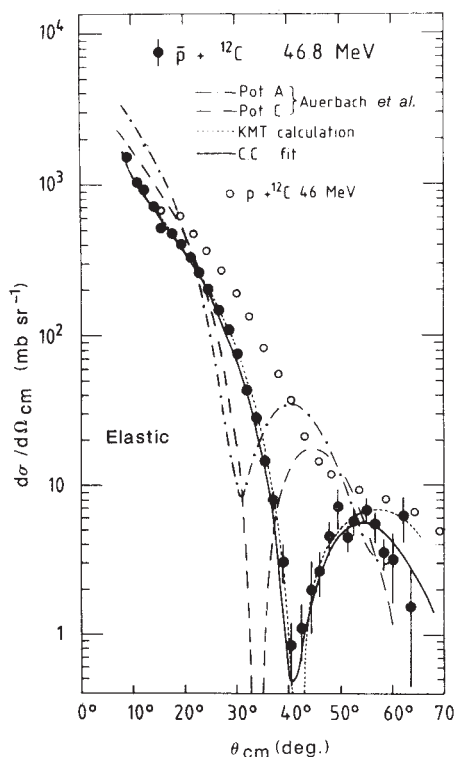


Fig. 3 Differential cross-sections for the elastic scattering of antiprotons at ^{12}C . For comparison the elastic scattering cross section of protons at ^{12}C is also given. The curves correspond to various calculations (ref. 20).

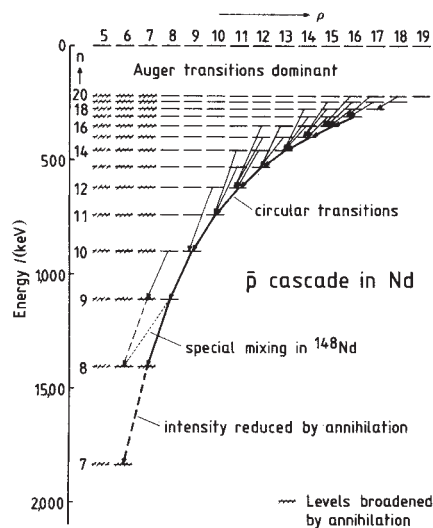


Fig. 4 Antiprotonic levels in the Nd atom. All measured X-ray transitions (see Fig. 5) are indicated by arrows. The E2-resonance effect causes special mixing of the transition (dashed) with nuclear excitation.

strong interaction attracts them in addition to the electromagnetic interaction, and they annihilate with a nucleon at the nuclear surface. The most important X-ray transitions indicated in the Nd scheme of Fig. 4 have been observed in the X-ray spectrum, measured using Ge detectors at LEAR (Fig. 5)²⁵. This demonstrates the beautiful correspondence established between the expected scheme, and the experimental results.

These experiments have been performed with stopped antiprotons. A beam of antiprotons from LEAR (21 or 47 MeV, between 10,000 and 100,000 s⁻¹) is moderated by a few mm of polyethylene in such a way that they stop in the target. The

target has an area of about 20 × 20 mm² and a thickness of a fraction of a mm. Several Ge detectors for γ and X rays and particle detectors view the target.

Impressive antiprotonic X-ray spectra of the noble gases He, Ne, Ar, Kr and Xe have been obtained with a cyclotron trap²⁶ which uses a magnetic field to concentrate antiprotons at low gas pressures (in these cases 50 mbar). These experiments prove that the antiproton depletes the electron shell by Auger processes during the cascade, if there is no refilling.

The strong interaction changes the energy and natural width of the last observable X-ray transitions. Therefore, these energy shifts ε (difference between the measured energies and the calculated electromagnetic energies, $\varepsilon = E_{\text{exp}} - E_{\text{em}}$) and broadenings Γ yield information on the antiproton-nucleus strong-interaction potential and annihilation. In Nd the last observable transition is ($n=8 \rightarrow n=7$). The intensity Y of the last observable transition is frequently reduced because a fraction of the antiprotons already annihilate in the upper level ($n=8$ for Nd). Thus the reduction of Y also gives information on the $\bar{p}$ absorption.

The strong interaction of antiprotons with protons or neutrons has two components. An interaction with longer range (≥ 0.8 fm) is given by the meson exchange ($\pi, 2\pi, \eta, \rho, \omega$). The short-range interaction is dominated by annihilation and quark degrees of freedom (see Fig. 1). A recent survey of this nucleon-antinucleon interaction can be found in ref. 27. As a first approximation for the strong interaction a complex antiproton-nucleus optical potential $\bar{V}_{\text{sl}}(r)$ is frequently added to the electromagnetic potential in the Schrödinger or Dirac equation. This optical potential is related to the complex average antiproton-nucleus scattering length $\bar{a}$ and the density $\rho(r)$ of the protons and neutrons²⁸:

$$\bar{V}_{\text{sl}}(r) = -\frac{2\pi}{m} \left(1 + \frac{m_{\bar{p}}}{m_N}\right) \bar{a} \rho(r) = V(r) + iW(r)$$

where m is the $\bar{p}$ -nucleus reduced mass, $m_{\bar{p}}$ the mass of the antiproton and m_N the nucleon mass. The potential can be separated into a real part, and an imaginary part that is responsible for the annihilation.

The complex scattering length $\bar{a}$ can be determined from the measured values of ε , Γ and Y . A fit to 18 nuclides measured before 1980 yielded²⁸

$$\bar{a} = (1.53 \pm 0.27) + i(2.50 \pm 0.25) \text{ fm}$$

However, new more precise measurements of antiprotonic atoms at LEAR gave (with similar errors for individual nuclei)

$$\bar{a} = (1.2 \pm 0.3) + i(2.4 \pm 0.3) \text{ fm}$$

for ^{16}O and ^{19}F (ref. 29) and

$$\bar{a} = (0.25 \pm 0.30) + i(3.05 \pm 0.25) \text{ fm}$$

for ^{92}Mo (ref. 30). The discrepancy between the new values indicates that either (i) the approximation with the optical potential is not adequate; (ii) the used nuclear charge density distribution does not describe the density correctly at the surface; (iii) $\bar{a}$ depends on the mass number A ; or (iv) a spin-orbit term has to be added to the potential. The spin-orbit term plays an essential role in the shell model of the nucleus and it has been questioned whether it should also be introduced in the antiproton-nucleus potential

$$\bar{V}(r) = V(r) + iW(r) + \text{ls}(V_{\text{so}}(r) + iW_{\text{so}}(r))$$

There is a strong indication from an experiment at LEAR³¹ that a small ls-term is also necessary for the antiproton absorption by nuclei.

As the annihilation takes place at the surface of the nucleus, all measurements with antiprotonic atoms only probe the outer range of the potential $\bar{V}(r)$. The nucleus looks essentially like a black sphere for the antiprotons. A simple model based on this idea³² of a totally absorbing nucleus, reproduces very nicely

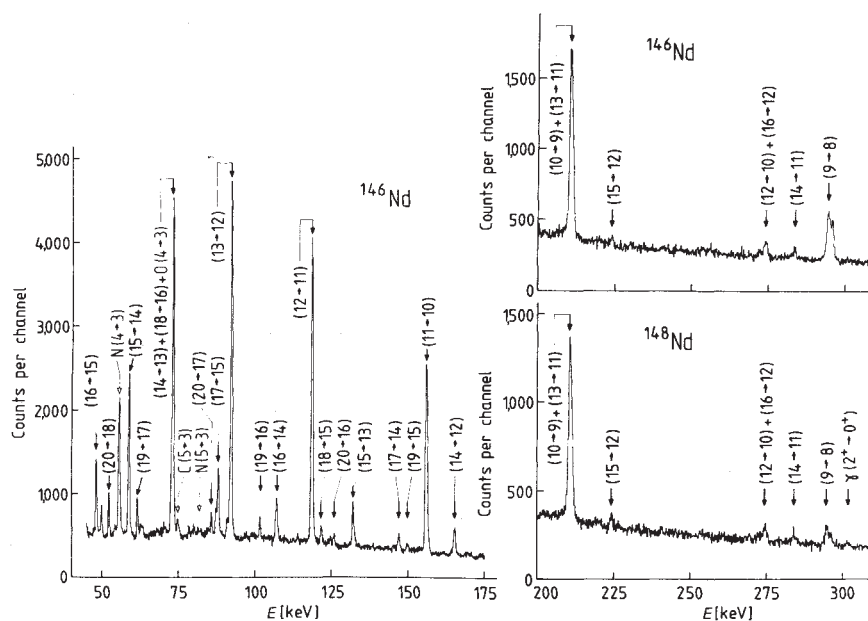


Fig. 5 Antiprotonic X-ray spectrum in Nd isotopes measured with Ge detectors. The numbers in brackets indicate the two principal quantum numbers n of the upper and lower level of the transition. The (9-8) transition in ^{148}Nd is reduced by the E2-resonance effect. The transitions with C, N or O are the background lines of the corresponding elements (refs 25, 31).

the antiprotonic data ϵ , Γ and Y , and in particular the new data measured at LEAR.

A particular phenomenon of antiprotonic atoms, is the E2 nuclear resonance effect, predicted by Leon³³ in 1975. This effect is a mixing of atomic and nuclear states wherever the spacing of antiprotonic levels is matched by the spacing of nuclear levels. The effect has been observed in antiprotonic Mo isotopes³⁰ and antiprotonic³¹ Nd by the attenuation of antiprotonic X-ray transitions as predicted. Some of these results have previously been discussed in *Nature*³².

Antiproton annihilation

The annihilation of slow antiprotons takes place at the nuclear surface with one proton or neutron. Annihilation with two nucleons may be possible¹⁵ but has not been observed yet.

The cross-section for the annihilation of a slow antiproton with a proton is about twice as large as that with a neutron³⁵. Calculations of the overlap of the nuclear wave function with the wave function of the antiproton in its last orbit show that the annihilation occurs preferentially in the outer surface of the nucleus where the nuclear density is only $\sim 10\%$ of the central density³⁶. In this region there is evidence for a neutron halo from antiprotons in heavier nuclei, for which annihilations with neutrons were found to be more frequent³⁷.

The combination of an antiproton with a proton or neutron during annihilation implies an enormous concentration of energy¹¹, 1,880 MeV in a space with $r = 1.9$ fm. It has been proposed^{11,14} that such a fireball melts nucleons to form a quark-gluon plasma by a phase transition. As mentioned above, experiments as yet have given no indications of such special states or large quark bags, hotspots, shock waves or density waves.

The annihilation is dominated by the emission of many pions which are frequently reached via intermediate mesonic resonances^{38,39}. Between two and eight pions (π^+ , π^- , π^0) are produced, that is on average five (three of them charged). The pion multiplicity distribution can be reproduced by statistical models⁴⁰. The fraction of pions that penetrate the nucleus have mostly a short range because their energy, ~ 100 – 300 MeV, is in the region of the Δ -resonance. Consequently much energy is transferred to the nucleus. Several groups have developed computer codes to simulate the intranuclear cascade^{14,36,41}, that is, the collisions of pions, deltas and nucleons and the emission of fast pions, protons and other particles. It has been calculated³⁶ that an energy of ≤ 800 MeV, on average 100–200 MeV, is deposited in the nucleus. This extremely hot nucleus sub-

sequently evaporates many neutrons and protons. A recent theoretical publication comes to the conclusion that significant amounts of thermal energy are not released within the nucleus during annihilation¹⁶. The energy spectra and the multiplicity of the emitted particles give information on the intranuclear cascade process and the energy distribution in the nucleus.

Particle emission

The study of particle emission following stopped antiproton annihilation is still in its early stages. Previous investigations⁴² used photographic emulsions for the identification of the charged particles. Experiments at LEAR also used photographic emulsions⁴³, a self-shunted streamer chamber in a magnetic field^{35,44}, a magnetic spectrometer system with many sophisticated detectors (CALLIOPE)⁴⁵, telescopes with Ge and Si detectors⁴⁶ and a multiplicity detector with 31 scintillation counters²⁵. Most of these instruments started with experiments which used antiprotons with energies between 20 and 200 MeV and measured cross-sections and charged-particle multiplicity distributions. Comparison with model predictions indicate that most annihilations occur at the surface, but some take place inside the nucleus with high multiplicities⁴³. Up to 13 charged particles were emitted after the annihilation of 20 MeV antiprotons in neon⁴⁷ and up to 18 charged particles after annihilation of stopped antiprotons in AgBr⁴³, with average values of 6.4 in neon and 4.9 in AgBr. The probability for the annihilation inside the nucleus increases with the antiproton energy.

The telescope measurements⁴⁶ using many targets from Li to U, yielded spectra of pions, protons (p), deuterons (d), tritons (t), ^3He and ^4He with energies up to 100 MeV. Even ^6He particles have been observed. The spectra have an exponential decrease towards higher kinetic energies E and can be fitted with the formula $C e^{-E/T}$. The parameter T is closely related to the average energy of the particles and seems to be approximately independent of the mass number A . This fact is also predicted by intranuclear cascade calculations¹⁴ and was observed in proton spectra emitted after high-energy proton collisions with nuclei⁴⁸. Preliminary evaluations of the measurements⁴⁶ indicate that the T parameter decreases with the mass of the emitted particle ranging from ~ 70 MeV for p to 15 MeV for ^4He . Surprisingly the conversion of these particle energies into momenta yields very similar values, about 350 MeV/c. The average momentum of the pions created in $\bar{p}p$ annihilation is 350 MeV/c. The ratios of the number of the emitted particles d/p, t/d, $^4\text{He}/^3\text{He}$ were determined⁴⁶ for many isotopes from Li to U. They range for d/p (with energies from ~ 20 MeV to ~ 180 MeV)

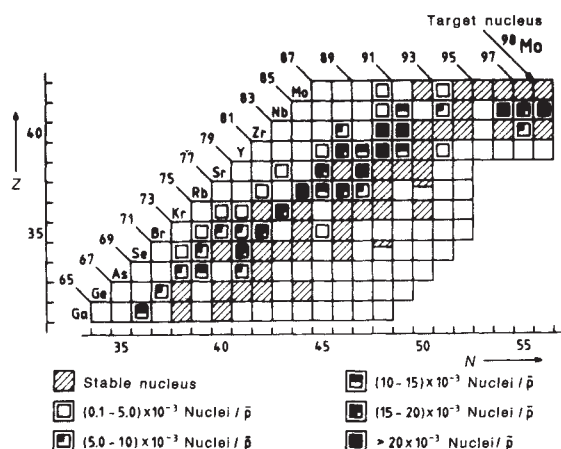
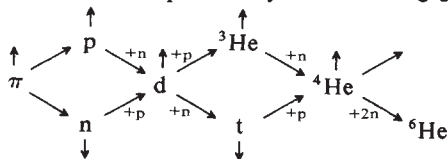


Fig. 6 Part of the chart of nuclides with the distribution of residual nuclei after $\bar{p}$ annihilation in ^{98}Mo . Observed residual nuclei are marked by squares (ref. 53).

from ~ 0.09 in ^7Li to ~ 0.2 in ^{232}Th and for $^4\text{He}/^3\text{He}$ (with energies from 40 to 50 MeV) from ~ 0.9 in ^{12}C to 2.4 in ^{238}U .

Assuming that the production of d , t , ^3He or ^4He is a consecutive process, a simple pickup theory might be able to reproduce, at least qualitatively, the experimental findings: the fast pions from the annihilation strike protons or neutrons which either escape or undergo multiple scattering picking up other nucleons and forming d , t , ^3He and ^4He . The average momentum (350 MeV/c) seems to be conserved. The consecutive and competing processes can be explained by the following graphical scheme:



The arrows indicate either escape (vertical arrows), or strikes by the pion, or pickup of a proton or neutron.

The probability of a striking pion π meeting a proton or neutron is proportional to Z/A or N/A , respectively (Z , proton number; N , neutron number of the nucleus). The pickup probabilities for $(d+p)$ are for instance given by $(Z/A)(1 - \exp(-r_d(E)A^{1/3}))$ or for $(^3\text{He}+n)$ by $(1 - \exp(-(N/A)r_{\text{He}}(E)A^{1/3}))$. The average pickup rates r (pickup probability per average distance in the nucleus) are independent of A as long as the Coulomb barrier does not play a role. The pickup probabilities have to be combined according to the scheme given above to obtain ratios of emitted particles. Most experimental particle ratios can then be reproduced using pickup rates between 0.05 and 0.3 per $A^{1/3}$.

The particle spectra and ratios are similar to the ones observed in fast pion-nucleus interactions^{49,50}, as expected.

Residual nuclei

What is left of a nucleus after annihilation? This important question has been investigated only very recently. Intranuclear cascade calculations from several groups^{15,36,41} predict the distribution of residual nuclei and can be tested by the new experimental results obtained at LEAR. A very broad distribution is expected because the multiplicity of the high-energy primary pions varies between two and eight and the number and strength of the interactions with the nucleus are quite different. Broad distributions have been observed after fast pion³¹ and fast proton⁵² reactions with nuclei. The distribution of residual nuclei can be estimated from the multiplicity distribution of charged particles measured, for instance, in nuclear emulsions.

The first measurements of yields of residual nuclei were performed at LEAR⁵³⁻⁵⁵. Metal sheets of ^{95}Mo , ^{98}Mo , ^{165}Ho and ^{238}U were irradiated for several hours each, with a total of about

2×10^8 antiprotons stopped in the targets. Starting a few minutes after $\bar{p}$ irradiation, the radioactivity was measured using Ge detectors for the gamma spectra. Thirty-three radioactive isotopes were identified in the ^{98}Mo target. Up to 30 nucleons were observed to be emitted from ^{98}Mo and up to 49 from ^{165}Ho . Figure 6 shows a part of the chart of the nuclides with the observed residual isotopes of the ^{98}Mo target. Only neutron-deficient isotopes were seen. In a few per cent of the annihilations only one proton and very few, if any, neutrons were removed from the target nucleus. In these cases evidently no pion entered the nucleus. The distribution of residual nuclei exhibits a substantial odd-even effect; yields of nuclides are reduced by 30% for unpaired protons and neutrons in the Mo isotopes and by 50% for unpaired protons in Ho only. This odd-even effect is probably caused by the fact that odd nucleons are less bound and evaporate more easily, producing even nucleon numbers. The experimental distributions are not yet well reproduced by the theoretical calculations⁴¹ because not enough energy is transferred to the nucleus in the calculations, so as to emit a sufficient number of particles.

The radioactivity after antiproton annihilation in ^{238}U stems essentially from neutron-rich nuclides with masses $\sim A = 110$ (ref. 55). This indicates that following the annihilation 10–20 nucleons are emitted, the hot nucleus fissions with a symmetric distribution, and the fission fragments evaporate additional nucleons.

Conclusion

This review has described the wealth of very interesting phenomena revealed in the interactions and annihilations of antiprotons and nuclei. Many open questions are stimulating further theoretical investigations. The antiproton-nucleus optical potential has to be improved and a spin-orbit term has to be introduced. The E2 resonance effect mixes atomic and nuclear states. Spectra of emitted particles and distributions of residual nuclei have been measured with several targets and are awaiting detailed theoretical interpretation. These experiments at LEAR have yielded the first survey of the dominant reactions between antiprotons and nuclei. The next step will be the search for rare reactions which have been predicted but not so far observed. All these experiments have only become possible using the very pure low-energy beam of antiprotons at LEAR, and CERN must be praised for the performance of its very sophisticated accelerator systems.

The interaction of antiprotons with nuclei is a simple example of the interaction of antimatter with matter. In this case enormous energies will be liberated and matter and antimatter will disintegrate into fast pions, fast nucleons and gamma rays. One kilogram of antimatter produces more than 2,000 times more energy than the fissioning of 1 kg of ^{235}U . Fortunately, the military applications of antimatter are completely out of reach. One antiproton produces about 2 GeV energy. But for the production of one stored antiproton one needs two million protons, each with 30 GeV. Therefore, at least 30 million times more electrical energy is needed to produce antimatter than is finally 'gained' by annihilation. Any contact between matter and antimatter leads to annihilation and destruction. Thus the storage of larger amounts of antimatter is impossible, as it would require neutralization with positrons, and probably very high magnetic fields or very low temperatures. A recent article⁵⁶ discusses the problems arising if low-energy antiprotons are stored.

Research with antiprotons and antimatter will continue to be a very interesting field of basic physics with no practical applications—and all questions of antimatter are closely related to the question of the formation of matter during the Big Bang.

I thank H. Daniel and F. J. Hartmann for many fruitful discussions and for critically reading this manuscript. The successful work of the LEAR staff and the friendly cooperation with many physicists at LEAR is gratefully acknowledged. I am indebted to P. Stoeckel for graphical assistance.

1. Dirac, P. *Proc. R. Soc. A* **126**, 360–365 (1930).
2. Dirac, P. *Proc. R. Soc. A* **133**, 60–72 (1931).
3. Anderson, C. D. *Phys. Rev.* **43**, 491–494 (1933).
4. Chamberlain, O., Segrè, E., Wiegand, C. & Ypsilantis, T. *Phys. Rev.* **100**, 947–950 (1955).
5. Cork, B., Lambertson, G. R., Piccioni, O. & Wenzel, W. A. *Phys. Rev.* **104**, 1193–1197 (1956).
6. Dorfán, D. E., Eades, J., Lederman, L. M., Lee, W. & Ting, C. C., *Phys. Rev. Lett.* **14**, 1003–1006 (1965).
7. Antipov, Yu. M. *et al. Nucl. Phys.* **B31**, 235–252 (1971).
8. Morrison, P. *Am. J. Phys.* **26**, 358–368 (1958).
9. Beverini, N. *et al. A Measurement of the Gravitational Acceleration of the Antiproton CERN Experiment PS200*, January 1986 (CERN, Geneva, 1986).
10. Goldman, T., Hughes, R. J. & Nieto, M. N. *Phys. Lett.* **171B**, 217–222 (1986).
11. Rafelski, J. *Phys. Lett.* **91B**, 281–284 (1980).
12. Strottman, D. *Phys. Lett.* **119B**, 39–42 (1982).
13. Cahay, M., Cugnon, J., Jasselette, P. & Vandermeulen, J. *Phys. Lett.* **115B**, 7–10 (1982).
14. Clover, M. R., DeVries, R. M., DiGiacomo, N. J. & Yariv, Y. *Phys. Rev. C* **26**, 2138–2151 (1982).
15. Cugnon, J. & Vandermeulen, J. *Phys. Lett.* **146B**, 16–20 (1984).
16. McGaughey, P. L., Clover, M. R. & DiGiacomo, N. J. *Phys. Lett.* **166B**, 264–266 (1986).
17. Bocquet, J. P. *et al. Phys. Lett.* **182B**, 146–150 (1986).
18. Billinge, R. *et al. in Physics with Antiprotons at LEAR in the ACOL Era* (eds Gastaldi, U. *et al.*) (Frontières, Gif-sur-Yvette, France, 1985).
19. Garett, D. *et al. in Physics with Antiprotons at LEAR in the ACOL Era* (eds Gastaldi, U. *et al.*) (Frontières, Gif-sur-Yvette, France, 1985).
20. Garett, D. *et al. Phys. Lett.* **135B**, 266–270 (1984); **139B**, 464 (1984).
21. Friedman, E. & Lichtenstadt, J. *Nucl. Phys. A* **455**, 573–583 (1986).
22. Wong, C.-Y., Kerman, A. K., Satchler, G. R. & MacKellar, A. D., *Phys. Rev. C* **29**, 574–580 (1984).
23. Auerbach, E. H., Dover, C. B. & Kahana, S. *Phys. Rev. Lett.* **46**, 702–705 (1981).
24. Kahana, S. H. & Sainio, M. E. *Phys. Lett.* **139B**, 231–234 (1984).
25. Kanert, W. thesis, Technische Univ., München (1986).
26. Bacher, R. *et al. Annual Report on Nuclear Physics Activities* 104–105 (KfK 4159 Kernforschungs-zentrum Karlsruhe, 1986).
27. Dover, C. B. *Nucl. Phys. A* **416**, 313c–334c (1984).
28. Batty, C. I. *Nucl. Phys. A* **372**, 418–432 and 433–444 (1981).
29. Poth, H. *et al. in Physics with Antiprotons at LEAR in the ACOL Era* (eds Gastaldi, U. *et al.* 581–591. (Frontières, Gif-sur-Yvette, France, 1985); *Phys. Lett.* **B176**, 327–333 (1986); *Z. Phys. A* **325**, 261–265 (1986).
30. Kanert, W. *Phys. Rev. Lett.* **56**, 2368–2371 (1986).
31. Daniel H. *et al. Z. Phys. A* **326**, 523–525 (1987).
32. Kaufmann, W. B. & Pilkuhn, H. *Phys. Lett.* **166B**, 279–281 (1986).
33. Leon, M. *Nucl. Phys. A* **260**, 461–473 (1976).
34. Batty, C. J. *Nature* **323**, 487 (1986).
35. Balestra, F. *et al. Nucl. Phys. A* **465**, 714–732 (1987).
36. Iljinov, A. S., Nazarov, V. I. & Chigrinov, S. E. *Nucl. Phys. A* **382**, 378–400 (1982).
37. Bugg, W. M., Condo, G. T., Hart, E. L., Cohn, H. O. & McCulloch, R. D. *Phys. Rev. Lett.* **31**, 475–478 (1973).
38. Armenteros, R. & French B. in *Antinucleon–Nucleon Interactions in High-Energy Physics* Vol. IV (ed. Burhop, E. H. S.) 237 (Academic, New York, 1969).
39. Baltay, C. *et al. Phys. Rev.* **145**, 1103–1111 (1966).
40. Orfanidis, S. J. & Rittenberg, V. *Nucl. Phys. B* **59**, 570–582 (1973).
41. Cugnon, J. & Vandermeulen, J. *Nucl. Phys. A* **445**, 717–756 (1985).
42. Ekspong, A. G., Frisk, A., Nilsson, S. & Ronne, B. E. *Nucl. Phys.* **22**, 353–409 (1961).
43. Batusov, Yu. A. *Europhys. Lett.* **2**, 115–122 (1986).
44. Balestra *et al. Nucl. Instrum. Meth.* **A234**, 30–41 (1985).
45. Sunier, J. W. *Nucl. Instrum. Meth.* **A241**, 139–152 (1985).
46. Daniel, H. *et al. Verhandl. DPG (VI)* **22**, IV, E-8.3 (1987).
47. Balestra, F. *et al. Nucl. Phys. A* **452**, 573–590 (1986).
48. Burgov, N. A. *et al. Soviet J. Nucl. Phys.* **24**, 620–622 (1976).
49. Amann, J. F. *et al. Phys. Rev. Lett.* **35**, 1066–1069 (1975).
50. Kaufman, S. B. *et al. Phys. Rev. C* **32**, 1977–1985 (1985).
51. Jackson, H. E. *et al. Phys. Rev. C* **18**, 2656–2665 (1978).
52. Cumming, J. B., Hausteijn, P. E., Stoenner, R. W., Mausner, L. & Naumann, R. A. *Phys. Rev. C* **10**, 739–755 (1974).
53. Moser, E. F. *et al. Phys. Lett.* **B179**, 25–29 (1986).
54. Moser, E. F. thesis, Technische Univ., München (1987).
55. Daniel, H. *et al. Verhandl. DPG (VI)* **22**, IV, C-2.6 (1987).
56. Maddox, J. *Nature* **324**, 299 (1986).

ARTICLES

Peculiar variations in the structure of the quasar 3C454.3

I. I. K. Pauliny-Toth*, R. W. Porcas*, J. A. Zensus†, K. I. Kellermann‡, S. Y. Wu§, G. D. Nicolson|| & F. Mantovani¶

* Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, D-5300 Bonn 1, FRG

† California Institute of Technology, Pasadena, California 91125, USA

‡ National Radio Astronomy Observatory, Edgemont Road, Charlottesville, Virginia 22901, USA

§ Beijing Observatory, Academia Sinica, Beijing, China

|| National Institute for Telecommunications Research, Johannesburg, RSA

¶ Istituto di Radioastronomia del CNR, Via Irnerio 46, Bologna, Italy

Following a major outburst at radio wavelengths in the quasar 3C454.3, observations by means of very long baseline interferometry have revealed a superluminal increase in the size of the compact 'core' region of the source and the appearance of a complex structure within it. Some features within this structure remained stationary with respect to each other, but others showed superluminal motion. This behaviour is distinctly different from that generally seen in superluminal sources.

THE radio source 3C454.3, one of the brightest extragalactic sources at centimetre wavelengths, is identified with a 16 mag quasar at a redshift $z = 0.859$. It is an optically violent variable and varies strongly at centimetre¹ as well as decimetre^{2,3} wavelengths. Its X-ray emission is weak; a single measurement with the Einstein observatory in 1980 gave a flux density of only 0.5 μ Jy (D. Harris, personal communication).

The time-scale of the variations at decimetre wavelengths led to the suggestion that bulk relativistic motion should be present⁴. Although the low-frequency variations are probably not intrinsic, but arise in the interstellar medium^{5,6}, the short time-scale and large amplitude of the variations even at centimetre wavelengths suggest that bulk relativistic motion is required to reduce the brightness temperature in the source frame below the limit set by inverse Compton losses⁷.

Two large outbursts at centimetre wavelengths have occurred in 3C454.3 within the last twenty years. The first occurred in 1967⁸, when very long baseline interferometry (VLBI) was in

its infancy, and only limited measurements were made following the outburst⁹. The second reached a peak at a wavelength of 2.8 cm in mid-1981 (Fig. 1), and we began VLBI measurements at that time.

Observations

Most of our VLBI observations were made at a wavelength of 2.8 cm with transatlantic baselines, giving an angular resolution of about 0.4 mas. The epoch of each measurement is indicated in Fig. 1, which also shows the variations in the flux density. We have also made observations at wavelengths of 6 cm and 18 cm, to map features of low surface brightness, and at 1.3 cm, to improve the angular resolution. Table 1 gives the wavelength, epoch and telescopes used for each observation. Preliminary results from some observations have already been reported^{10,11}.

At all wavelengths longer than 1.3 cm, the data were recorded using the Mark II recording system¹², which has an effective bandwidth of 1.8 MHz, and were correlated with the three-

Table 1 VLBI observations of 3C454.3

λ (cm)	Epoch	Telescopes
1.3	1982.5	SKGOY
	1983.8	BSKGOY
	1985.8	BSKGOY
2.8	1981.4	BKGFO
	1982.1	BKGFO
	1983.8	BKGFHO
	1984.1	BKGF(H)O
	1984.6	BKGFHO
	1984.9	BLKGFHO
	1985.3	B(LK)GFHO
	1985.6	BL(K)GF(H)O
	1985.9	(L)KG(F)HO
	1986.2	(B)LKG(F)HO
6	1981.6	BKGFO
	1983.9	BWJSGFOE
18	1982.9	RSBJDwFGO

The codes for the telescopes, their diameters and locations are as follows. B: 100 m, Effelsberg, FRG. Dw: 25 m, Dwingeloo, Netherlands. E: 25 m, Hartebeesthoek, RSA. F: 25 m, George Agassiz Station, Ft. Davis, Texas, USA. G: 43 m, NRAO, Green Bank, West Virginia, USA. H: 25 m, Hat Creek, California, USA. J: 76 m, Jodrell Bank, Cheshire, UK. K: 37 m, Haystack Observatory, Westford, Massachusetts, USA. L: 32 m, Medicina, Italy. O: 40 m, Owens Valley Radio Observatory, California, USA. R: 22 m, Simeiz, Crimea, USSR. S: 26 m or 20 m, Onsala Space Observatory, Onsala, Sweden. W: 43–50 m (3 to 4 antennae of array), Westerbork, Netherlands. Y: 25 m (one antenna of array), NRAO, Socorro, New Mexico, USA. Parentheses around the code for a telescope mean a technical failure so that most, or all of the data from that antenna were lost.

station Mark II processor at the Max-Planck-Institut für Radioastronomie (MPIfR) in Bonn. At a wavelength of 1.3 cm, observations with the Mark II system at epochs 1982.5 and 1983.8 gave poor results owing to insufficient sensitivity. They were repeated at epoch 1985.8 using the Mark III recording system¹³ with a bandwidth of 28 MHz and gave satisfactory results. These latter observations were correlated with the MPIfR four-station Mark III processor.

To improve the data quality at 2.8 cm at epochs after 1983.8, when the source became heavily resolved and the correlated amplitudes decreased, we used a new fringe-fitting algorithm¹⁴ following the correlation. This procedure, which uses antenna-based residuals in fringe-rate and delay, significantly improved the quality of the maps. The calibration of the data followed the procedures of Cohen *et al.*¹⁵ and the subsequent production of maps used the usual steps of Fourier inversion, CLEAN and self-calibration¹⁶ with software developed by J. Romney and W. Alef at MPIfR. To obtain quantitative results for the strengths and separations of features in some of the maps, components with Gaussian brightness distributions were fitted to the visibility data.

Results

Source structure. The structure of 3C454.3 is shown in Fig. 2 over a wide range of angular scales. On the arcsec scale, the MERLIN map (Fig. 2a), made at a wavelength of 75 cm with a resolution of about 1 arc s¹⁷, shows a strong core feature, and a weaker 'jet', pointing towards a 'hotspot' some 5 arc s away at position angle -48° . Our VLBI observations at 18 cm wavelength, epoch 1982.9, which have a resolution of 2.5 marc s, show a similar structure (Fig. 2b), but on a scale some 250 times smaller; the jet here is only about 20 marc s long (180 pc; we adopt a Hubble constant $H_0 = 55 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $q_0 = 0.05$ throughout), and its position angle varies from -65° near the core to -53° at its end. There is no evidence for a 'counter-jet', down to a level of 2% of the brightness of the main jet. The map in Fig. 2b is very similar to earlier maps obtained at 18 cm

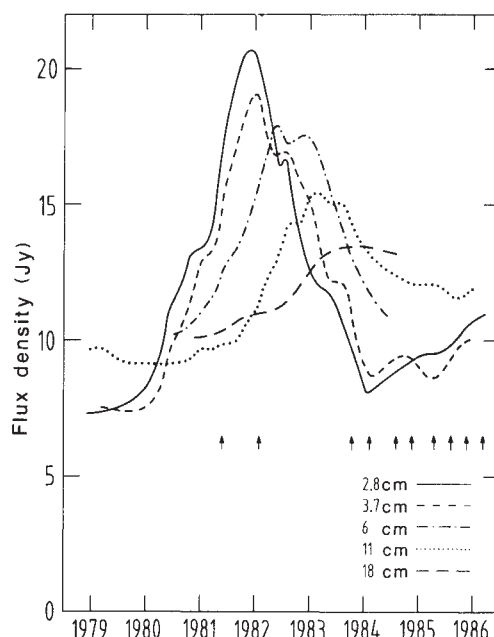


Fig. 1 Variations in the total flux density of 3C454.3 at five wavelengths, shown schematically as smooth curves drawn from the original data. These data are from our own observations at wavelengths of 6 and 18 cm, from our own observations and J. M. MacLeod, P. A. Feldman and B. H. Andrew, personal communication, at 2.8 cm and from ref. 31 at 3.7 and 11 cm. The epochs of our VLBI monitoring observations at 2.8 cm are shown by arrows.

wavelength, at epochs 1980.1 and 1981.8, which had a higher north-south resolution¹⁸. The structure has remained unchanged at this wavelength between 1980.1 and 1982.9.

The inner part of the jet is seen in more detail in the map made at 6 cm wavelength, epoch 1983.9 (Fig. 2c) with a resolution of 1.0 marc s. At this wavelength, the jet can be traced to about 10 marc s (90 pc) from the core. At 2.8 cm wavelength, the jet is difficult to map, owing to its low surface brightness and the limited dynamic range of the maps. We show in Fig. 2d a map made at 2.8 cm at epoch 1984.6, when the peak brightness of the core was relatively low, so that the jet can be seen. At this wavelength and epoch, the core region was well resolved and consisted of at least three features, extending over 1.7 marc s (15 pc) at position angle -95° . At 1.3 cm, epoch 1985.8 (Fig. 2e) the core region is extended in the same position angle, and the easternmost feature is the brightest. The position angle thus changes through an overall angle of about 45° over a range of linear scales from a few pc to 50 kpc, a property characteristic of other asymmetric, core-dominated objects such as the superluminal source 3C345 (ref. 19).

Radio spectrum. Near epoch 1981.5, VLBI maps of 3C454.3 are available over a range of wavelengths from 2.8 to 18 cm. We have used these data to derive the radio spectra of the core region and of the milliarcsecond jet separately, as shown in Fig. 3. The core region had a spectrum characteristic of a self-absorbed source at wavelengths longer than 2.8 cm, with a spectral index $\alpha \sim 1.1$ (flux density $S \propto \nu^\alpha$), indicating that the core region is not homogeneous. The milliarcsecond jet, on the other hand, has a spectrum typical of a transparent source, with $\alpha \sim -0.8$.

We note that at a wavelength of 75 cm, where variations of the total flux density of up to 4 Jy are observed²⁰, the extrapolated flux density of the core region is only about 1 Jy, while that of the milliarcsecond jet is about 10 Jy. This suggests that at long wavelengths the variations arise from the jet feature, and, in view of its large extent, are presumably extrinsic, as suggested by Rickett *et al.*⁵.

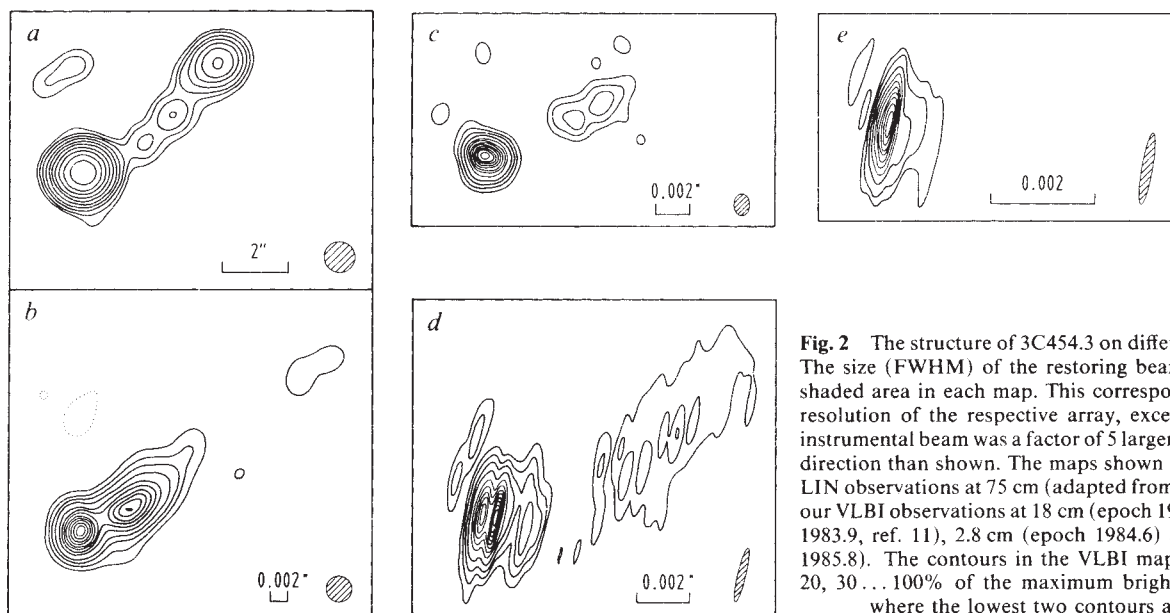


Fig. 2 The structure of 3C454.3 on different angular scales. The size (FWHM) of the restoring beam is shown by the shaded area in each map. This corresponds roughly to the resolution of the respective array, except in *b*, where the instrumental beam was a factor of 5 larger in the north-south direction than shown. The maps shown are: *a*, from MERLIN observations at 75 cm (adapted from ref. 17); *b*–*e*, from our VLBI observations at 18 cm (epoch 1982.9), 6 cm (epoch 1983.9, ref. 11), 2.8 cm (epoch 1984.6) and 1.3 cm (epoch 1985.8). The contours in the VLBI maps are 1, 2.5, 5, 10, 20, 30...100% of the maximum brightness, except in *e*, where the lowest two contours are omitted.

Structural changes. We show the results of our VLBI monitoring observations at 2.8 cm wavelength in Fig. 4, both as a series of maps of the core region and as corresponding cuts through the maps along position angle -95° , the direction in which the region is elongated. The structural changes seen are quite unlike those which have been observed in the 'classical' superluminal sources, such as 3C120, 3C179, 3C273, or 3C345 (ref. 21).

The first two sets of observations, at epochs 1981.4 and 1982.1, near the peak of the outburst at 2.8 cm wavelength, show the core as only slightly resolved, with no evidence for any components having an overall separation of more than 0.3 marc s. At both epochs, the data are well fitted by a model consisting of an unresolved component ≤ 0.2 marc s (1.8 pc) and a larger component, about 0.5 marc s (5 pc) in diameter, centred on the

same position. Between 1981.4 and 1982.1, the total flux density of the core region increased by about 2 Jy and this increase occurred in the *larger* component, which brightened by about 50%. The brightening of a region 5 pc in size, in a period of only 0.7 years may be described as 'superluminal', with an apparent propagation rate of the exciting signal (if originating at the centre) $\geq 20c$.

By epoch 1983.8, the structure had changed dramatically; the core region became elongated at position angle -95° and showed three distinct brightness peaks, with an overall separation between the extreme peaks of 1 marc s (9 pc). There is an indication of a further component 1.4 marc s west of the easternmost feature. We refer to these features by number, starting from the east. If these features were ejected from a region < 0.3 marc s in extent at epoch 1982.1, in one direction, the apparent velocities of components 2, 3 and 4 between 1982.1 and 1983.8 would have to have been greater than 10c, 28c and 40c, respectively. These values are consistent with that implied by the 'superluminal brightening' observed between 1981.4 and 1982.1.

To follow the rapid motions implied by these data, we made frequent observations from early 1984 on. As well as the maps and cuts in Fig. 4, we show, in Fig. 5, the separations of the components from component 1, and their peak brightness, plotted against the epoch. In both figures, we have assumed that component 1 is the 'true', stationary core.

Surprisingly, following the rapid increase in size of the core region between 1982.1 and 1983.8, the separations of the brightest features remained nearly constant. With respect to component 1, the separation of component 2 changed by less than 0.05 marc s yr^{-1} (apparent velocity $< 3c$), while that of component 3 showed no significant change during the limited period when it was visible.

Nevertheless, component 4, which became distinctly visible on the western side of the core region by 1984.1, appears to have shown 'classical' superluminal motion between 1983.8 and 1984.9, when it separated from component 1 at a rate of 0.35 ± 0.06 marc s yr^{-1} (apparent velocity $19 \pm 3c$). A weaker, diffuse feature (component 5?) even further west had a similar rate of separation (0.21 ± 0.05 marc s yr^{-1} , or $12 \pm 3c$). After 1984.9, these outer features faded, and the presence of any motions became difficult to establish.

This fading has occurred in all components with the exception of 1: component 3 had disappeared by 1984.9, and even component 2 has become steadily fainter since 1984.9. Component

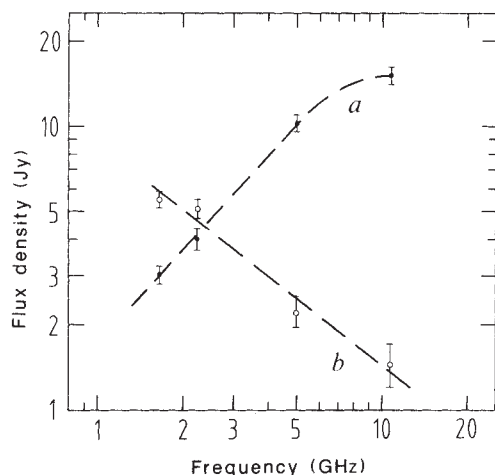


Fig. 3 The radio spectrum of the core region (*a*) and the milli-arcsecond jet (*b*) of 3C454.3 at epoch 1981.5. The flux densities plotted were obtained by integration of our own VLBI maps at 2.8 and 6 cm (10.7 and 5 GHz) near this epoch, at 18 cm (1.65 GHz) by interpolation of the results of Padrielli (ref. 18) obtained at epochs 1980.1 and 1981.8, and at 13 cm (2.29 GHz) from the map of Cotton *et al.*³². The latter was based on VLBI measurements at epoch 1980.6, and we have assumed that the increase in flux density between 1980.6 and 1981.5 can be attributed to an increase in the core flux density. The increase was derived from data at 11 cm (Fig. 1) and scaled to 13 cm using the spectrum of the core region between 18 cm and 6 cm. The correction was 0.6 Jy.

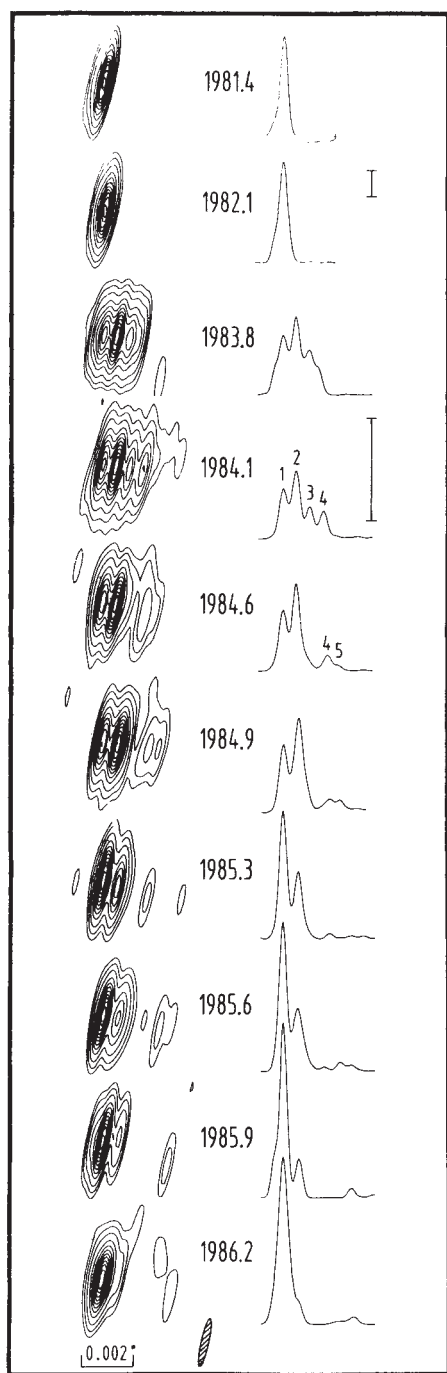


Fig. 4 Structural variations in the core region of 3C454.3. On the left are shown hybrid maps for 10 epochs, obtained from our VLBI monitoring programme at 2.8 cm. On the right, on the same angular scale, are shown cuts through the brightness distributions in the direction of extension of the core region (position angle -95°). Both maps and cuts have been obtained with the same restoring beam of 2×0.3 mas (shown shaded). The actual instrumental beam was, on the average, 2.4×0.4 mas, with the exception of epoch 1985.9, when the absence of European telescopes gave a beam about two times larger. The contour levels in the maps are 2.5, 5, 10, 20, 30... 100% of the peak brightness temperatures of 19, 19, 3.8, 3.3, 4.1, 4.4, 6.3, 7.2, 8.4 and 8.1×10^{10} K, in order of epoch. The brightness scale for the cuts is shown by the two vertical bars on the right, which represent a brightness temperature of 4.8×10^{10} K for the first two epochs, and the remaining epochs, respectively. The features referred to in the text are labelled in the cuts for 1984.1 and 1984.6. Since hybrid maps contain no absolute position information, the maps and cuts have been aligned on the assumption (justified in the text), that component 1 is the stationary core and that this is the only feature seen at the first two epochs.

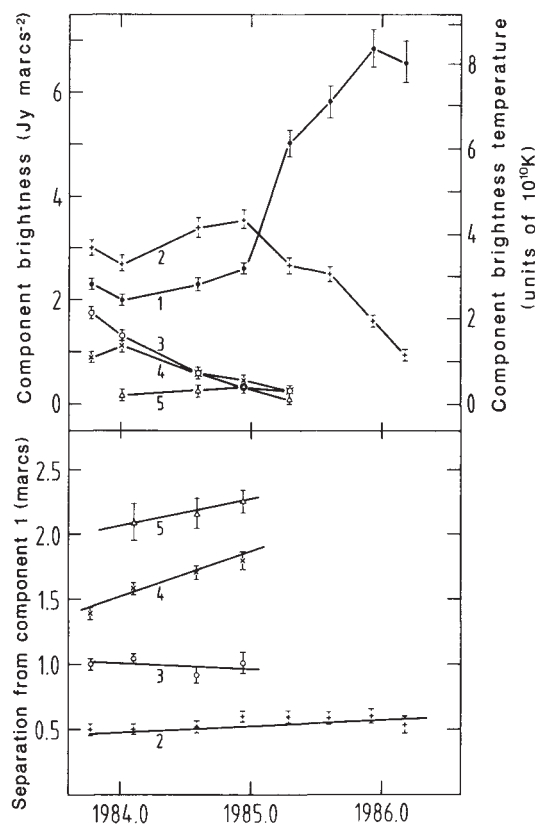


Fig. 5 Variations in the brightness (top) and separations (bottom) of the components in the core region of 3C454.3. The values plotted, and their errors, have been obtained by fitting the appropriate number of Gaussian profiles to the data plotted as cuts in Fig. 4. Note that the brightness for the first two epochs (not plotted) was about $16 \text{ Jy (mas s)}^{-2}$ (1.9×10^{11} K), and we attribute this to component 1 (see text). The separations, in mas are referred to this component. The straight lines in the bottom plot are weighted least-squares fits to the data points and have slopes of 0.045 ± 0.016 , -0.046 ± 0.072 , 0.347 ± 0.058 and $0.210 \pm 0.045 \text{ mas yr}^{-1}$ for components 2, 3, 4 and 5, respectively. The corresponding apparent velocities are $(2.5 \pm 0.9)c$, $-(2.6 \pm 4.0)c$, $(19.2 \pm 3.2)c$ and $(11.6 \pm 2.5)c$, respectively.

1, however, has brightened steadily since 1984.1, when the total flux density at 2.8 cm wavelength also began to increase again. The brightening of component 1 (which is the brightest feature at 1.3 cm wavelength) and the superluminal brightening between the first two epochs suggest, first, that component 1 is the true core and, second, that the feature observed at the first two epochs also corresponds to this component, so that the alignment of the maps and cuts in Fig. 4 is correct.

Interpretation

The structural changes observed in the core region of 3C454.3 do not appear to fit easily into the scheme of relativistic jet models which have been used successfully to account for the observations of the 'classical' superluminal sources. The rapid variations in flux density, the high radio luminosity and the large transverse motions observed suggest that bulk relativistic effects, of the kind inferred for other quasars with similar properties, are also important in 3C454.3. The structural variations in 3C454.3, however, appear to be much more complex than in the 'classical' superluminal sources. The observations show a rapid increase in size of the core region (with apparent velocity $v_a \geq 30c$) and the appearance of some features which are nearly stationary with respect to component 1, while others move with

$v_a \sim 20c$. To explain this complex behaviour, we discuss several interpretations: (1) the rapid increase in size was not real, but was due to fading of the brightest feature, which caused previously present, but undetected, features to appear above the threshold of the map. This explanation is not tenable, as there was a significant increase in emission outside the brightest feature, from less than 1 Jy at epoch 1982.1, to about 6 Jy at 1983.8.

(2) The rapid increase in size resulted from a true motion of distinct components. On the standard models²² this requires relativistic bulk motion with Lorentz factor $\gamma \geq 30$ and an angle between the motion and the line of sight, $\theta \leq 2^\circ$. The presence of features showing no relative motion after 1983.8 could then be due to (i) a decrease in the bulk velocity at some time between 1982.1 and 1983.8. For $\gamma = 30$, $\theta = 2^\circ$, a change in velocity of $\sim 3000 \text{ km s}^{-1}$ is required to reduce v_a from $30c$ to $3c$. Although this is a modest change in velocity, it is not clear why it should occur for two features (2 and 3) at different distances from the core. Moreover, the corresponding change in energy ($\Delta E/E = \Delta\gamma/\gamma \sim 75\%$) is substantial. (ii) A change in the direction of motion. For the same initial conditions ($\gamma = 30$, $\theta = 2^\circ$), a change of θ to about 0.1° (or 38°) would reduce v_a to $3c$. If the bend in the trajectory occurs at the same place for all components, however, the observed separations 1-2, 1-3 are not explained, since the change in θ causes components to 'pile-up' at the bend. On the other hand, if the bend occurs at a larger distance from the core for 3 than for 2, then identical amounts and directions of bending are required, which seems unlikely. (iii) A misidentification of the true core. If the true, stationary core is invisible, either because it is always opaque at cm wavelengths, or because the beam axis at the core makes a large angle with the line of sight, a group of features moving at the same rate with respect to it would appear stationary. The rapid increase in size would then correspond to the emergence of successive features from the opaque region, or their motion past a suitable bend in the trajectory (say to $\theta \sim 2^\circ$). This explanation is also not tenable, since 3C454.3 shows emission from a compact region, presumably the core, even between outbursts. (3) The features in the core region may not represent discrete, physical components, but rather regions where the emission from a relativistic jet is enhanced, for example by shocks. Lind and Blandford²³ have emphasized that the velocity of shock fronts, which might be observed as a motion of 'components', can be

smaller than that of the underlying medium, which determines the Doppler boosting of the emission. In particular, it is possible to obtain stationary emission patterns.

We believe that this is the most likely explanation for the complex behaviour of 3C454.3. Following the outburst, the relativistic fluid excites shocks at progressively larger distances from the core. The resulting appearance of emission features is superluminal: the separation of features 2 and 4, which appeared sometime between 1982.1 and 1983.8, requires an apparent velocity for the fluid of at least $30c$, corresponding to $\gamma > 30$, $\theta < 2^\circ$. Near the core, the shock pattern appears stationary. A change in physical conditions at larger distances from the core produces a moving pattern and gives rise to the 'classical' superluminal behaviour of the outer components (4 and 5).

Although the 'classical' superluminal sources (for example, 3C120, 3C179, 3C273, 3C345) do not show features that are stationary with respect to their cores, 3C454.3 is not unique in this respect. In the quasar 4C39.25, a pattern of two strong components, with a constant separation, was observed over several years^{24,25}, but more recent observations have revealed the appearance of a weaker, superluminal feature moving between the relatively stationary components²⁶⁻²⁸. A similar phenomenon has been reported in the quasar 3C395 (ref. 29). The radio galaxy 3C111 is reported to show a behaviour similar to that of 3C454.3 after the initial outburst: following a period during which no motions with respect to the core were observed, all the features faded, leaving only the core³⁰. In 3C454.3 itself, a period during which an extended feature faded rapidly has also been reported⁹.

The present work demonstrates that simple models (such as ref. 22) are inadequate in explaining all the structural changes observed in superluminal sources. Although the observed phenomena fit into the general framework of relativistic beaming models, the increased complexity which now appears to be required means that the models are less able to predict the behaviour of the sources and are correspondingly less attractive or useful.

We thank Drs J. MacLeod and R. Fiedler for providing data on the variability of 3C454.3 before publication. Thanks are also due to the staff at the various observatories, and at the MPIfR processor. The National Radio Astronomy Observatory is operated by Associated Universities, Inc., under contract with the National Science Foundation.

Received 18 May; accepted 3 July 1987.

- Kellermann, K. I. & Pauliny-Toth, I. I. K. *Nature* **213**, 977-980 (1967).
- Hunstead, R. W. *Astrophys. J. Lett.* **12**, 193-200 (1972).
- Fanti, R., Ficarra, A., Mantovani, F., Padrielli, L. & Weiler, K. *Astr. Astrophys. Suppl.* **36**, 359-364 (1978).
- Jones, T. W. & Burbidge, G. R. *Astrophys. J.* **186**, 791-799 (1973).
- Rickett, B. J., Coles, W. A. & Bourgois, G. *Astr. Astrophys.* **134**, 390-395 (1984).
- Rickett, B. J. *Astrophys. J.* **307**, 564-574 (1986).
- Kellermann, K. I. & Pauliny-Toth, I. I. K. *Astrophys. J. Lett.* **155**, L71-L78 (1969).
- Kellermann, K. I. & Pauliny-Toth, I. I. K. *A. Rev. Astr. Astrophys.* **19**, 373-410 (1981).
- Gubbay, J. *et al. Astrophys. J.* **215**, 20-35 (1977).
- Pauliny-Toth, I. I. K., Porcas, R. W., Zensus, J. A. & Kellermann, K. I. *IAU Symp.* **110**, 149-152 (Reidel, Dordrecht, 1984).
- Wu, S. Y., Pauliny-Toth, I. I. K., Porcas, R. W. & Zensus, J. A. *Acta astrophys. sinica* **6**, 89-100 (1986).
- Clark, B. G. *Proc. IEEE* **61**, 1242-1248 (1973).
- Rogers, A. E. E. *et al. Science*, **219**, 51-54 (1983).
- Alef, W. & Porcas, R. W. *Astr. Astrophys.* **168**, 365-368 (1986).
- Cohen, M. H. *et al. Astrophys. J.* **201**, 249-255 (1975).

- Pearson, T. J. & Readhead, A. C. S. *A. Rev. Astr. Astrophys.* **22**, 97-130 (1984).
- Browne, I. W. A., Orr, M. J. L., Davis, R. J., Foley, A., Muxlow, T. W. B. & Thomasson, P. *Mon. Not. Roy. astr. Soc.* **198**, 673-688 (1982).
- Padrielli, L. *et al. Astr. Astrophys.* **165**, 53-73 (1986).
- Browne, I. W. A. *et al. Nature* **299**, 788-793 (1982).
- Padrielli, L. *IAU Symp.* **110**, 169-176 (Reidel, Dordrecht, 1984).
- Cohen, M. H. & Unwin, S. W. *IAU Symp.* **110**, 95-103 (Reidel, Dordrecht, 1984).
- Blandford, R. D. & Königl, A. *Astrophys. J.* **232**, 34-48 (1979).
- Lind, K. R. & Blandford, R. D. *Astrophys. J.* **295**, 358-367 (1985).
- Shaffer, D. B. *et al. Astrophys. J.* **218**, 353-360 (1977).
- Bäth, L. B. *et al. Astrophys. J.* **232**, 34-48 (1979).
- Marcaide, J. M. *et al. Nature* **314**, 424 (1985).
- Shaffer, D. B. & Marscher, A. P. *Bull. Am. Astr. Soc.* **17**, 609 (1985).
- Shaffer, D. B., Marscher, A. P., Marcaide, J. M. & Ronney, J. D. *Astrophys. J. Lett.* **314**, L1-L5 (1987).
- Waak, J. A., Spencer, J. H., Johnston, K. J. & Simon, R. S. *Astr. J.* **90**, 1989-1991 (1985).
- Götz, M., Alef, W., Preuss, E. & Kellermann, K. I. *Astr. Astrophys.* (in the press).
- Fiedler, R. L. *et al. Astrophys. J. Suppl.* (in the press).
- Cotton, W. D. *et al. Astrophys. J.* **286**, 503-508 (1984).

Limits on hadronic cosmic ray production by young pulsars

David Eichler* & John R. Letaw†

* Department of Physics and Astronomy, University of Maryland, College Park, Maryland 20742, USA and Department of Physics, Ben Gurion University, POB 653, Beer Sheva 84105, Israel.

† Severn Communications Corporation, Box 544, Severna Park, Maryland 21146, USA

Since the discovery of supernova 1987A (Shelton) several authors^{1,2} have noted that it may provide an excellent opportunity to observe the cosmic ray output of a young pulsar through the 'beam dump' that the supernova ejecta provide. It has been suggested³ that neutrino emission from p-p collisions is possible immediately, and that ultra-high energy γ -ray emission might be possible after several months, when the supernova remnant becomes transparent to them. In this letter we argue that the cosmic abundances of Li, Be and B set significant constraints on the cosmic ray proton production in the young ($t \leq 1$ yr) remnant, and, in particular, rule out neutrinos from shock-accelerated protons in the ejecta at currently detectable levels.

In a typical Type II supernova ($10\text{--}15 M_{\odot}$), the inner parts of the ejecta are believed to be the C and O burning shells⁴. Approximately $1 M_{\odot}$ of heavy elements ($Z \geq 6$) is typically ejected. Most of the interstellar C and much of the O is thought to be synthesized in these events. In this case, cosmic rays (CR) produced by the young pulsar within the first several months interact with and fragment this highly enriched material, producing the light elements Li, Be and B. The C/O layer remains thick to cosmic rays for at least $5 \times 10^6 (M_{C+O}/M_{\odot})^{1/2} u_3^{-1}$ s if the CR are untrapped by magnetic fields, and longer if trapped. (Here M_{C+O} is the combined mass of ejected C and O, and u_3 is the ejection velocity divided by $3 \times 10^3 \text{ km s}^{-1}$).

The relative branching ratios for production of light elements in proton interactions with ^{12}C , ^{14}N , ^{16}O and ^{20}Ne are given in Table 1, together with their cosmic abundances relative to ^{12}C .

Table 1 Branching ratios at $E \geq 300$ MeV per nucleon and interstellar abundances of light elements Li, Be and B

Product	Cosmic abundance* ($^{12}\text{C} \equiv 1$)	Targets			
		^{12}C	^{14}N	^{16}O	^{20}Ne
		1.0	0.21	1.64	0.21
Branching ratios† for p on target:					
^6Li	4×10^{-7}	0.06	0.06	0.05	0.05
^7Li	5×10^{-6}	0.08	0.08	0.07	0.06
^9Be	1×10^{-7}	0.03	0.03	0.02	0.03
^{10}B	2×10^{-7}	0.09	0.08	0.04	0.03
^{11}B	6×10^{-7}	0.22	0.10	0.08	0.05

* Ref. 5.

† Refs 6, 7 and 8.

Protons passing through much more than 84 g cm^{-2} of C (one mean free path) will almost always interact producing, for example, ^{10}B in 9% of the reactions.

Defining ζ to be the total number of nuclear collisions per primary in the hadronic cascade, we estimate ζ to be $9 \ln_2 (\epsilon/\text{GeV})$, where ϵ is the primary energy, by the following argument: ultrarelativistic protons lose half their energy per collision, each generating a knock-on nucleon of several GeV. The total number of ultrarelativistic collisions and mildly relativistic knock-ons are thus each $\sim \ln_2 (\epsilon/\text{GeV})$. Mildly relativistic protons collide more elastically and go through ~ 3 generations of collisions before being degraded to the ionization loss-domi-

nated regime, $E \leq 150$ MeV, and we assume each collision yields two protons which share the collision energy. Thus each proton starting at 1 to 2 GeV causes $\sim 2^3$ mildly relativistic collisions. Including the original primary, we arrive at $9 \ln_2 (\epsilon/\text{GeV})$ total collisions per primary.

If $10^{50} E_{50}$ ergs in hadronic CR are produced by the pulsar within the first month or so, then the total number of them is $10^{53} E_{50}/1.6(\bar{\epsilon}/\text{GeV})$, where $\bar{\epsilon}$ is their average energy. The total number of, say, ^{12}C nuclei is $1 \times 10^{56} (M_C/M_{\odot})$, so if the protons traverse pure carbon the predicted $^{10}\text{B}/\text{C}$ number ratio is $5.6\zeta \times 10^{-5} (E_{50} M_{\odot}/(\bar{\epsilon}/\text{GeV}) M_C)$, which should be compared with the observed cosmic ratio of 2×10^{-7} . The quantity $E_{50}/(\bar{\epsilon}/\text{GeV})$ must therefore be less than about $3.6/\zeta \times 10^{-3}$. Comparable limits come from the $^6\text{Li}/\text{C}$ ratio and from oxygen and (to a lesser extent) neon spallation.

At a distance of 55 kpc, neutrinos from 10^{50} ergs in primary CR protons produce ~ 1 neutrino interaction per kiloton of detecting material. This implies that if very-high energy (VHE) neutrinos are detectable by kiloton detectors from the supernova in the Large Magellanic Cloud, they must have a very hard spectrum, $\bar{\epsilon} \geq 30$ TeV.

For muon neutrinos, the effective detector mass is determined by the range of the daughter muon and the area. Detailed calculations¹ indicate that 10^{49} ergs in 10 TeV primaries yields ~ 1 count/ 10^2 m^2 . Hence, according to the above arguments, $\bar{\epsilon}$ cannot be much less than 3 TeV if the muon neutrinos are to be detectable by present day detectors.

Neutrinos from shock-accelerated primaries, for which $\bar{\epsilon}$ is typically only several GeV, appear to be limited by the above considerations to below currently detectable levels. Monochromatic particle beams from the pulsar itself could easily have an $\bar{\epsilon}$ in excess of 3 TeV, but it is unclear that they would have large hadronic components. If the more easily detectable VHE γ -rays are eventually observed and prove to indicate a much higher $E_{50}/\bar{\epsilon}$, it will be evidence that the pulsar is putting out mainly e^+e^- pairs as opposed to hadronic cosmic rays.

We now consider the impact of this limit on possible γ -ray line emissions. Protons generated in the hadronic cascade within a thick C shell and degraded to 100 MeV or less yield 4.44 MeV γ -rays from direct excitation reactions. The cross section for this reaction ranges from 300 mb at 10 MeV to 10 mb at 100 MeV⁹. Assuming that (i) 50% of the cascade protons interact in the outer 40 g cm^{-2} of the shell where the photons can escape and (ii) 20% of these interactions lead to 4.44 MeV line emission, the line intensity from a source at 55 kpc is $\sim 3 \times 10^{-3} E_{50}(\zeta/\bar{\epsilon})$ photons $\text{cm}^{-2} \text{ s}^{-1}$. The limit on high-energy protons therefore implies that fewer than 10^{-5} photons $\text{cm}^{-2} \text{ s}^{-1}$ will be observed from line emissions. This flux is not detectable by current instruments (SMM) and marginally detectable by the Gamma-Ray Observatory (GRO).

The above argument if applied to pulsars in general might be somewhat weakened by the possibility that most carbon may come from supernovae that do not leave rapid pulsars. However supernova 1987A appears to be a type II supernova of typical mass¹⁰, and the hadronic CR content of the shell is then constrained to the extent that it is typical. The C, O and Ne in the remnant may become transparent to cosmic rays, at $t \geq 1$ month, before the remnant as a whole does, so that the heavy elements are in effect diluted with H and He, and the resulting ratios of Li or B to C, O or Ne are diminished by the dilution factor. Alternatively, there may be mixing during the explosion. However, these effects weaken the limit on $E_{50}/\bar{\epsilon}$ by at most an order of magnitude or so.

The basic argument can be generalized to any explosive process in the galaxy. The cosmic abundance of ^{10}B is readily explained¹¹ by the spallation of cosmic rays (at their present level) of interstellar C and O, and processes that result in more spallation are ruled out if the spallation products find their way back into the interstellar medium. Scenarios for greatly enhanced

$\bar{p}$ production^{12,13}, are possible only if the CR escape into the ISM without the irradiated beam dump material. Neutron stars, if they are to account for $\bar{p}$ excess, must also serve as the waste-dump sites for irradiated material.

This research was supported in part by NSF. We are grateful to Drs T. Vestrand and J. Goodman for helpful discussions.

Received 22 May 1987; accepted 17 July 1987.

1. Gaisser, T. K. & Stanev, T. *Phys. Rev. Lett.* **58**, 1695-1697 (1987).
2. Sato, H. (preprint).
3. Bresinsky, V. S. & Prilutsky, O. F. *Astr. Astrophys.* **66**, 325-334 (1978).
4. Woosley, S. E. & Weaver, T. A. *A. Rev. Astr. Astrophys.* **24**, 205-253 (1986).
5. Cameron, A. W. in *Essays in Nuclear Astrophysics* (eds Barnes, C. A., Schramm, D. N. & Clayton, D. D.) Vol. 23 (Cambridge University Press, 1982).
6. Olson, D. L., Berman, B. L., Greiner, D. E., Heckman, H. H., Lindstrom, P. J. & Crawford, H. J. *Phys. Rev. C* **28**, 1602-1613 (1983).
7. Letaw, J. R., Silberberg, R. & Tsao, C. H. *Astrophys. J. Suppl.* **51**, 271-276 (1983).
8. Silberberg, R. & Tsao, C. H. *Astrophys. J. Suppl.* **25**, 315-368 (1973).
9. Ramaty, R., Koslovsky, B. & Lingenfelter, R. E. *Astrophys. J. Suppl.* **40**, 487-526 (1979).
10. Arnett, D. W. (preprint). Univ. of Chicago.
11. Walker, T. P., Mathews, G. T. & Viola, V. E. *Astrophys. J.* **299**, 745-751 (1985).
12. Cowdik, R. & Gaisser, T. K. in *Proc. 17th Int. Cosmic Ray Conf. (Paris)* Vol. OG, 3.1-3.16 (1981).
13. Eichler, D. *Nature* **295**, 391-393 (1982).

Statistical significance of the relationship between X-ray luminosity and variability timescale in active galactic nuclei

J. B. Stephen, L. Bassani, E. Caroli & G. Di Cocco

Istituto TESRE, CNR, Via De' Castagnoli 1, 40126 Bologna, Italy

Recently, Barr and Mushotzky¹ have claimed that a significant correlation exists between the X-ray luminosity and the timescale of X-ray variability for Seyfert galaxies and quasars. In particular, they find that $\log L_x = a_1 \log \Delta t + b_1$, with $a_1 = 0.9 \pm 0.2$, $b_1 = 39 \pm 1$ and linear correlation coefficient $r = 0.7$ (hereafter fit 1), as would be expected if these objects have a universal efficiency of conversion of matter into energy ($\eta \geq 0.005$, ref. 2). Furthermore, this result indicates that the emitting plasma is close to the Eddington luminosity limit, where electron-positron pair effects may dominate the hard X-ray photon fluxes³. In view of the physical meaning implied by their linear relationship and of its possible widespread use in the literature as a way of deriving constraints on the masses of active galactic nuclei^{4,5}, the method of analysis and its statistical significance should be examined carefully. In particular, because in this case both variables, L_x and Δt , are affected by experimental errors, their least squares fitting procedure is inaccurate and a more general approach must be taken. Applying this more generalized fit to the same data set, we obtain a slope $a = 1.5 \pm 0.2$, significantly different from that found by Barr and Mushotzky and more acceptable in terms of goodness of the fit. When judged on the basis of their statistical significance, both results are, in any case, too weak to be conclusive, and any physical conclusions drawn from them are therefore debatable.

In using the least squares procedure for fitting a line to a set of observational data the usual method is to assume that the experimental errors are all in, or attributable to, one variable—the dependent variable—and that the other parameter is known to infinite precision. Barr and Mushotzky have assumed that the errors in both variables can be concentrated in the luminosity L_x , and that the variability timescale Δt is measured exactly. The choice of the dependent variable is governed by the relative sizes of the errors: in general, the parameter with the largest error should be chosen. Because the quoted errors on L_x and Δt are of similar dimensions (see Fig. 1 in ref. 1), there is no *a priori* reason for choosing either as the dependent variable. In fact, exchanging the role of L_x and Δt and using the same data set, we find that for Seyfert galaxies and quasars $\log \Delta t =$

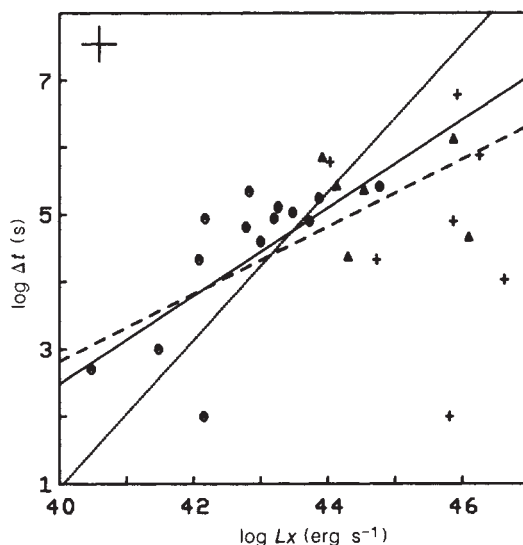


Fig. 1 Plot of $\log L_x$ against $\log \Delta t$ for the Barr and Mushotzky sample of Seyfert galaxies (circles), quasars (triangles) and BL Lac objects (crosses). A typical error bar is shown in the upper left-hand corner¹. Best fit lines to the data (excluding all BL Lac objects) refer to: all errors concentrated on $\log L_x$ (dotted line; fit 1 in the text), on $\log \Delta t$ (dashed line; fit 2 in the text) and equal errors on both variables (full line).

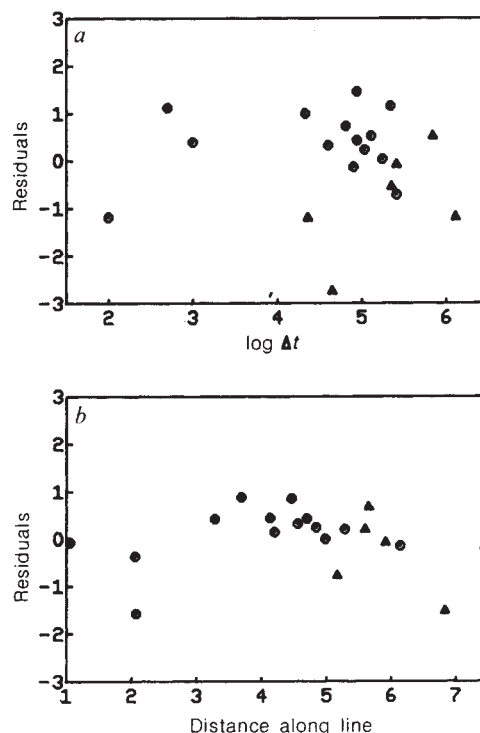


Fig. 2 Plot of the residuals calculated in the particular case of all the errors concentrated on one variable, $\log L_x$ (a) and in the case of equal errors on both variables, $\log L_x$ and $\log \Delta t$ (b) (taking in each case only the data relative to Seyfert galaxies and quasars). In b, the best fit line itself is used as the x-axis for the purposes of the plot.

$a_2 \log L_x + b_2$, with $a_2 = 0.5 \pm 0.13$ and $b_2 = -17.2 \pm 5.63$ (hereafter fit 2). This corresponds to $\log L_x = 2 \log \Delta t + 34.4$. In Fig. 1, both fits to the data (excluding the BL Lac objects) are plotted for comparison: the difference between the two straight lines is evident and significant. It is interesting to note that the linear correlation coefficient is symmetric with respect to the choice of the dependent variable and therefore cannot be used

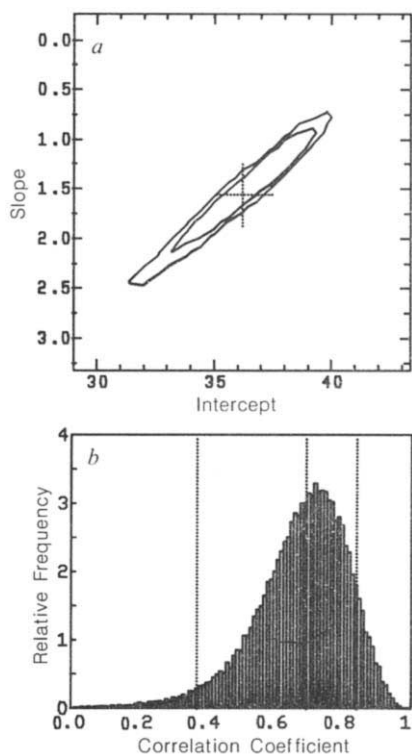


Fig. 3 *a*, Plot of the 68% and 90% confidence contours for the best fit parameters *a* (slope) and *b* (intercept) in the case of 100,000 bootstrap samples; the location of the point which corresponds to the best fit values for *a* and *b* calculated employing the more general fitting technique is indicated by a cross. *b*, Histogram of the frequency distribution of the correlation coefficient *r*; the dashed lines indicate the 90% confidence level on *r*.

to decide if either line is the more appropriate fit to the data. A method to determine whether either or both fits are consistent with the data and the quoted errors is the χ^2 -test. Applied to both fits, this test gives reduced $\chi^2(1) = 9.7 \pm 0.3$ and $\chi^2(2) = 7.8 \pm 0.3$ (for 18 degrees of freedom, d.f.), where the variance used is given by the variance of the dependent variable plus the variance of the independent variable multiplied by the square of the slope *a*. (On the other hand if we assume no *a priori* knowledge of the relationship between the two variables, that is, of the slope, and just sum the individual variances, the corresponding values of χ^2 are 8.8 and 4.9.) These values suggest that either both lines are bad fits, the errors are grossly underestimated or both. But if we assume that a straight-line relationship between the two variables does exist, fit 2 requires a smaller adjustment of the errors, indicating that perhaps it is a better fit to the data. In any case, both variables are affected by experimental errors and, therefore, the observed data point might be considered as having expanded to an ellipse, the ratio of whose axes is the ratio of the errors on the variables⁶⁻⁸. In this more general case, the best straight line is the one fitted in such a way that the sum of the squares of the distances from the centres of the ellipses to the line, measured along some appropriate direction is a minimum. The appropriate direction depends upon the relative errors of the two observed variables and is, in most cases, an oblique direction. In the particular case of variables having equal experimental errors, as indeed assumed by Barr and Mushotzky in their plot of $\log \Delta t$ against $\log L_x$, the appropriate direction is the perpendicular. The best fit line calculated in this way is plotted as a full line in fig. 1 for comparison with the same data set and with fits 1 and 2. For Seyfert galaxies and quasars, we find that $\log L_x = a \log \Delta t + b$, with $a = 1.53 \pm 0.16$, $b = 36.2 \pm 0.2$ and reduced $\chi^2 = 3.6 \pm 0.3$ (18 d.f.), where the variance used to evaluate the χ^2 is the sum of the two individual variances. The residuals for fit

1 and for the more general case are shown in Fig. 2*a, b*. It is evident that in Fig. 2*b* the objects are dispersed over a smaller range than in Fig. 2*a*.

To test the statistical reliability of the least squares fit parameters *a* and *b*, we have applied the bootstrap resampling analysis to the set of data under investigation^{9,10}. Unlike the χ^2 statistic, the bootstrap does not require an explicit error estimate associated with each data point; rather the errors are considered through the distribution of the data themselves. We have drawn 100,000 bootstrap samples from the original set of 20 data pairs and have applied to each sample the more general fitting technique discussed above. The results are presented in Fig. 3*a, b* where we have plotted the 68%, and 90% confidence contours for the best fit parameters *a* (slope) and *b* (intercept) as well as the corresponding frequency distribution of the correlation coefficient *r*. It is evident from Fig. 3 that the slope can range from 0.7 to 2.5 at the 90% confidence level and that the corresponding confidence limits for the correlation coefficient *r* are 0.35–0.83 (consequently the probability of achieving this by chance from a random sample ranges approximately from $0.1 \cdot 10^{-5}$).

The X-ray luminosity and variability timescale appear to be correlated, but the data available do not tend to support a linear relationship between $\log L_x$ and $\log \Delta t$. However, if the associated errors are considerably larger than assumed, a straight line may well be a good fit, but in this case our analysis suggests that a slope of ~ 1.5 is more appropriate than a slope of ~ 1 . Under these circumstances, it is unwise to draw any physical conclusions until more accurate observations are available.

The authors thank Dr S. Cecchini for useful discussions and constructive comments on the manuscript.

Received 1 December 1986; accepted 30 April 1987.

1. Barr, P. & Mushotzky, R. F. *Nature* **320**, 421–423 (1985).
2. Fabian A. C. *Proc. R. Soc.* **366**, 449–453 (1979).
3. White N. E., Fabian, A. C. & Mushotzky R. F. *Astr. Astrophys.* **133**, L9–L11 (1983).
4. Worrall, D. M. *Proc. Workshop The Continuum Emission in Active Galactic Nuclei*, January 1986, Tucson, Arizona (in the press).
5. Elvis, M., Czerny, B. & Wilkes, B. J. *Proc. ESA Workshop The Physics of Accretion onto Compact Objects*, April 1986, Tenerife, Spain (in the press).
6. Mack, C. *Essentials of Statistics for Scientist and Technologists* Ch. 12, 106–115 (Plenum, New York, 1967).
7. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* Chs 6, 7, 92–133, 187–203 (McGraw-Hill, New York, 1969).
8. Beers, Y. *Introduction to the Theory of Errors* Ch. 6, 38–43 (Addison-Wesley, Reading, Massachusetts, 1962).
9. Efron, B. *SIAM Rev.* **21**, 460–480 (1979).
10. Simpson, G. & Mayer-Hasselwander, H. *Astr. Astrophys.* **162**, 340–348 (1986).

BARR AND MUSHOTZKY REPLY—We would like to thank Stephen *et al.* for testing the statistical robustness of our result. We feel that the bootstrap analysis presented by the authors strengthens our result. We have considered the points raised by Stephen, Bassani, Caroli and Di Cocco and would like to reply to some of them.

Stephen *et al.* are correct in that the regression analysis we performed assumed that all of the error lay in the luminosity and that the timescale was well measured. This was because much of the data were derived from High Energy Astrophysical Observation satellite (HEAO-1) and Ariel-V measurements where the variability amplitude is large and thus well determined while the mean count rate is subject to background subtraction uncertainties and fundamental limits due to fluctuations in the diffuse X-ray background and therefore is less well determined. Another way of expressing this is that while the change in luminosity can be well determined the absolute counting rate has a relatively large systematic error.

Perhaps our quoted typical error bar was misleading as it showed symmetrical errors, we apologize for this. In reality, for most of our objects the true error is in the luminosity, L , is larger than the error in Δt . However, the errors are probably not normally distributed and in some sense are 'systematic'. In particular our derivation of the doubling time is subject to error for those objects with small amplitude variability if the true

shape of the light curve deviates strongly from our exponential assumption.

The reverse correlation slope s' is not just $s' = 1/s$. As shown in Bevington¹ the slopes are related by $s = R/s'$ where R is the regression coefficient. Stephen *et al.* seem to be unaware of this.

In the case where the errors are systematic the χ^2 statistic is not necessarily correct because the underlying assumption of gaussian distributed errors is not justified. In addition, as the errors on each data point have not been given and can differ significantly from the quoted 'representative' error bar we feel that the χ^2 analysis used by Stephen *et al.* is not valid. χ^2 per degree of freedom is large, the authors find a value of 3.6, the $\chi^2 + n$ contours are poor estimates of the true probability (as shown by Lampton *et al.*²). When χ^2 per degree of freedom is large the true errors are likely to be larger than determined by the $\chi^2 + n$ contour, but you cannot calculate them explicitly. As pointed out by Simpson and Mayer-Hasselwander³ this is often a reason for doing the bootstrap analysis.

The χ^2 statistic tests whether a detailed model is a good fit to the data while the regression analysis shows what the most likely fit to the data is, that is, shows the trend. Thus the fact that χ^2 is large indicates that the model is not a good description of the data but does not indicate that the data are not correlated. Therefore the regression analysis is not invalidated just because χ^2 is large.

We feel that the bootstrap analysis confirms our result. Our quoted best fit and error bars are consistent with the bootstrap analysis. But we are unsure what Stephen *et al.* mean by "the confidence limits on the correlation coefficient are 0.35–0.83". We interpret their analysis to be in agreement with our paper⁴.

The χ^2 analysis could show (if the true measurement errors were used) that the observed dispersion from the best fit cannot be explained simply by measurement error. We feel that some fraction of this could be due to the large uncertainty in the bolometric correction factor that we used and the simplicity of the power law model.

If our model is correct the correlation should be interpreted in the light of the total luminosity of the object and we simply used the 2–10 keV luminosity as an indicator of the total. Using the observed dispersions in power law indices of ± 0.15 and of high energy cutoffs in the spectra (0.1–3 MeV) this introduces an uncertainty in the bolometric correction of 0.5 in the logarithm.

It is also likely that the assumption of a constant Eddington ratio over such a large dynamic range in luminosity is incorrect. It is already clear that the most luminous quasars with $L_{\text{bol}} \sim 10^{48} \text{ erg s}^{-1}$ must be radiating at $> 10^{-3}$ of Eddington otherwise the mass of the central object would exceed the mass of a galaxy. It is thus likely that at the high end of our sampled luminosity range the assumption of a constant Eddington ratio, for instance, a power-law model for Δt against L , is incorrect.

In conclusion, the work of Stephen *et al.* does not appear to support their own conclusion in the abstract that "both results are, in any case, too weak to be conclusive". In fact the last paragraph of the text says that "the X-ray luminosity and variability timescale appear to be correlated" which is consistent with our original paper⁵. The bootstrap analysis is a new and original statistical validation of our original conclusions.

P. BARR

Exosat Observatory, European Space Operations Center,
Robert Bosch Strasse 5, 6100 Darmstadt, FRG

R. MUSHOTZKY

Laboratory for High Energy Astrophysics, NASA Goddard Space
Flight Center, Greenbelt, Maryland 20771, USA

Increased transition temperatures in $\text{YBa}_2\text{Cu}_3\text{O}_x$ superconducting ceramics by exposure to nitrogen

D. N. Matthews, A. Bailey, R. A. Vaile, G. J. Russell
& K. N. R. Taylor

Advanced Electronic Materials Technology Group,
School of Physics, University of New South Wales, PO Box 1,
Kensington, NSW 2033, Australia

The discovery of metallic oxide ceramics with superconducting transition temperatures in excess of 77 K (refs 1, 2) has led to an explosive growth of research on these materials. These solids are based on the oxygen-deficient orthorhombic perovskite^{3,4} $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, in which changes in the value of δ significantly affect the superconducting properties. These changes can be produced by annealing in various environments or by the detailed processing technique during sample manufacture (see, for example, ref. 5). In addition to the effects of oxygen concentration, it is known that the materials are susceptible to atmospheric attack through the combined effects of H_2O and CO_2 absorption. These effects lead to the development of a carbon-rich, insulating surface layer⁶, and subsequent loss of superconducting behaviour. Such degradation effects require that materials be kept in inert conditions at all times. During a recent series of observations of the transport properties of these materials, we discovered that exposure to helium gas at low temperatures causes dramatic changes in the nature of the transition⁷. Here we report the results of a parallel study using nitrogen gas, in which we observed increases of almost 40 K in the transition temperature in a nitrogen environment.

Specimens of $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ were prepared using our standard powder metallurgy route⁸ from dry powders of Y_2O_3 , BaCO_3 and CuO . Stoichiometric amounts were thoroughly mixed by grinding for 30 min before pelletizing at 10^4 Pa with a small amount of alcohol as wetting agent. These mixtures were then heated in air to 950°C at $\sim 500^\circ\text{C h}^{-1}$ and sintered at this temperature for 3 h. After cooling to room temperature, regrinding and repelletizing, the sintering process was repeated for a further 3 h at 950°C , followed by a slow cool to 550°C in oxygen at atmospheric pressure where they were left to anneal for 6 h. The sample was then cooled slowly to room temperature at 100°C h^{-1} . X-ray powder diffraction showed the sample to have the pattern characteristic of the orthorhombic phase⁹, with no evidence of a second phase.

Electrical resistivity measurements were made using a conventional four-probe technique in an Oxford Instruments continuous-flow cryostat. Sample dimensions were approximately $15 \times 4 \times 1 \text{ mm}$ and copper electrodes were attached with 'Dolite' silver paste.

Figure 1a shows the temperature variation of the resistivity for the sample cooled in vacuum to 77 K and heated to 110 K at $\sim 0.3 \text{ K min}^{-1}$. This sample displays a low-resistivity pedestal, reaching zero resistivity at 88 K, but with the midpoint of the major transition at 98.8 K. After this run the sample was immediately cooled from 110 to 77 K, using nitrogen as an exchange gas, and the measurement was repeated under the same conditions. The form of the $\rho(T)$ variation was essentially identical, with a slight displacement (0.5 K) to higher temperatures as shown in Fig. 1b.

The specimen was then left for 36 h at atmospheric pressure in nitrogen. Twenty of these hours were at 77 K, and during the following 16 h its temperature rose to room temperature. Re-cooling to 77 K in nitrogen led to the behaviour shown in Fig. 1c, with a relatively sharp transition, with little or no pedestal, and a transition temperature of 111 K. After a further 3 h in nitrogen at 77 K the zero-resistance temperature increased to 125 K (Fig. 1d), and then to 131 K 12 h later (Fig. 1e). The highest temperature reached between successive runs is shown in Fig. 1 by the vertical bars.

1. Bevington, P. R. *Data Reduction and Error Analysis for the Physical Sciences* Chs 6, 7, 92–133, 187–203 (McGraw-Hill, New York, 1969).

2. Lampton, M., Margon, B. & Boyer, S. *Astrophys. J.* **208**, 177 (1976).

3. Simpson, G. & Mayer-Hasselwander, H. *Astr. Astrophys.* **162**, 340–348 (1986).

4. Barr, P. & Mushotzky, R. F. *Nature* **320**, 421–423 (1985).

5. Mushotzky, R. F., Holt, S. S. & Serlemitsos, P. *Astrophys. J.* **225**, L115–L118 (1978).

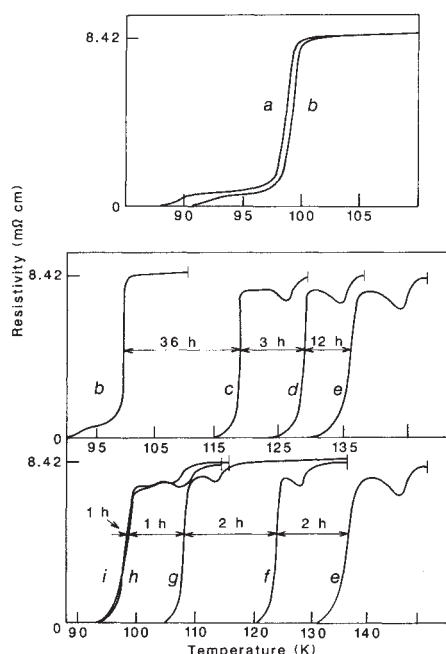


Fig. 1 Variation of resistivity with temperature of a $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ sample during continuous exposure to nitrogen gas at low temperatures. The top panel shows the starting conditions (*a* in vacuum; *b* in N_2 exchange gas), the middle panel shows the increase of transition temperature with continuing nitrogen addition, and the bottom panel shows the rapid decline of the peak value after 131 K had been reached. Run designations (*a*–*i*) refer to the sequential series of observations. Vertical bars on the curves show the highest temperature reached during the run, before cooling to 77 K for the next cycle.

A small resistivity anomaly appeared in run *c* at temperatures near the transition onset. This is similar to effects seen in a helium environment⁷, and its form remained constant during runs *c*–*e*.

Continued measurements in identical conditions resulted in a rapid decrease of the transition temperature, as shown in Fig. 1*e*–*i*. After repeated cycling in a nitrogen atmosphere the transition temperature stabilized at 92 K, slightly below the midpoint of the major transition in the original sample.

To establish the general nature of this nitrogen intrusion effect, $\rho(T)$ observations were made on two other samples prepared with lower pelletization pressures which had been stored in nitrogen gas at room temperature for about two weeks. The original transition temperatures of these samples had been 97 K; after nitrogen storage the zero-resistance temperatures were 106 and 108 K.

Finally, a complete cycling experiment was carried out for another sample prepared by our normal route. This gave identical results to those shown in Fig. 1. The upper limit for the transition temperature was 126 K in this case, but again exposure to a nitrogen atmosphere beyond this time resulted in a rapid return to the initial transition temperature.

We conclude that changes of >30 K can be induced in the transition temperatures of these superconductors by careful control of the nitrogen content at 77 K. This increase is significantly greater than that of 15 K obtained with helium gas⁷, and does not suffer from the serious distortion of the transition region. It is important to be aware of this type of effect in comparing measurements on different materials, or those of different research groups. These materials are quickly and easily affected by environmental gases in a variety of ways and we suggest any observational results should be accompanied by a detailed pedigree.

In trying to understand the origin of these changes, it is

tempting to assume that nitrogen enters the orthorhombic lattice along the relatively open channels running between the O–Cu–O chains in the *a*- or *b*-directions as occurs during nitrogen annealing at elevated temperatures (D. R. Clarke, personal communication), where it has been suggested that the nitrogen atoms occupy oxygen sites in the structure.

If this is so, then the associated bonding changes must lead to significant variations in the electron energy states at the Fermi level and consequently to changes in the strength of the superconducting interaction. It is also possible that the changes in the lattice parameters caused by the presence of the nitrogen atoms leads to a shift in the transition temperature equivalent to that produced by external pressure. To verify this model, it will be necessary to carry out a detailed structure determination of the material; this will be done during the coming weeks.

The sudden decrease in transition temperature must be caused by a change in the nature of the site being filled by the nitrogen atoms at a concentration corresponding to the highest transition temperature of ~131 K. It is well known that the addition of excess oxygen to the lattice can result in degradation of superconducting properties⁵, an effect which is thought to be due to the occupation of the normally vacant oxygen sites in the *a*–*b* plane that lie between the copper–oxygen chains. If nitrogen can enter oxygen sites in the lattice, then it is possible that the decrease in transition temperature results from the occupation of these same sites.

Received 23 July; accepted 7 August 1987.

1. Wu, M. K. *et al.* *Phys. Rev. Lett.* **58**, 908–911 (1987).
2. Zhao, Z. X. *et al.* *Kexue Tongbao* **32**, 412–414 (1987).
3. David, W. F. *et al.* *Nature* **327**, 310–312 (1987).
4. Welch, D. O. *et al.* *Nature* **327**, 278–279 (1987).
5. Gopalakrishnan, J. K. *et al.* *Nature* **327**, 604–605 (1987).
6. Russell, G. J. & Taylor, K. N. R. *Aust. Physicist* **24**, 140–141 (1987).
7. Matthews, D. N. *et al.* *Solid St. Commun.* (submitted).
8. Sun, H. B. *et al.* *Aust. Physicist* **24**, 138–139 (1987).
9. Cava, R. J. *et al.* *Phys. Rev. Lett.* **58**, 1676–1678 (1987).

Relationships between concentration and deposition of nitrate and sulphate in precipitation

G. Farmer*, R. J. Barthelmie*, T. D. Davies*, P. Brimblecombe† & P. M. Kelly*

* Climatic Research Unit and † School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

It has been shown that annual NO_3^- concentrations in Swedish precipitation rose from the late 1950s to the early 1970s and then remained constant at around the early 1970s level¹. The mean concentration in precipitation of a chemical species is one measure of pollution but of possible greater relevance for potential ecological damage is deposition, the total amount of the species input to the surface. We demonstrate here, using data from Eskdalemuir, Scotland (55°19' N, 3°12' W), that time trends in sulphate and nitrate deposition may differ in response to the influence of changes in precipitation amount. The consequence is that a change in precipitation amount alone will produce changes in the relative sulphate and nitrate deposition levels. Moreover, such changes could occur even if other parameters such as relative emission rates remained constant. Consequently, any assessment of nitrate/sulphate deposition should consider climatological as well as emission factors.

The European Air Chemistry Network (EACN) database used here contains, amongst other data, monthly values for precipitation and nitrate and sulphate concentration^{2,3}. As in ref. 1, extreme outliers in the monthly chemistry data were removed before analysis. Excluding values greater than four standard

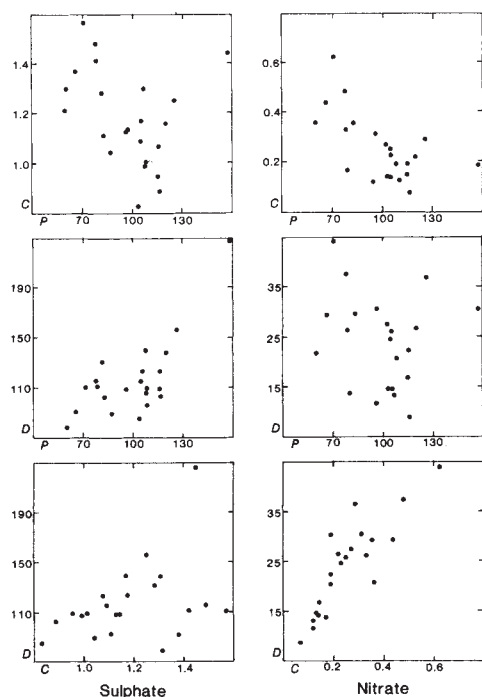


Fig. 1 Eskdalemuir EACN annual data 1958-79, mean monthly values for the year. Scattergrams of precipitation (P) and concentration (C) and deposition (D) for NO_3^- and excess SO_4^{2-} . Mean monthly values are plotted to compensate for those years which contain months with missing data. Rainfall units are mm, concentration units mg l^{-1} and deposition units mg m^{-2} . SO_4^{2-} as S; NO_3^- as N.

deviations from the mean eliminates one SO_4^{2-} concentration value (August 1970). A very high dry deposition component is suspected with precipitation of 6 mm and an uncorrected SO_4^{2-} value of 18 mg S l^{-1} . The sulphate data (in mg S l^{-1}) were then corrected for the proportion attributable to sea salt by subtraction of the factor 0.084 of the sodium concentration⁴. Eskdalemuir has previously been chosen as having useful data on the basis of a 'long data series, good ionic balance and no known local contamination'³. For both nitrate and sulphate, the annual mean concentration (C) values discussed here are precipitation-weighted averages of the monthly concentration (c) and monthly precipitation (p) data, that is:

$$C = \frac{\sum_{i=1}^{12} p_i c_i}{\sum_{i=1}^{12} p_i}$$

Annual deposition (D) values are given by

$$D = \sum_{i=1}^{12} p_i c_i$$

Annual precipitation-weighted concentration data may not reflect deposition levels because they ignore the variability in precipitation totals.

Relationships between precipitation and concentration mean that relationships between concentration and deposition are complex. Using the EACN annual data (1958-79) for Eskdalemuir, Fig. 1 shows the relationships between precipitation (P), concentration (C) and deposition (D) for both nitrate and sulphate. There is a negative correlation between P and C in the case of both nitrate and sulphate. For nitrate and to a lesser extent sulphate, a log-log relationship gives an improved fit. Similar relationships are present when the monthly values for P and C are plotted for nitrate and sulphate. This confirms that the correlations are not spurious nor a function of long-term trends. However, it is the $P:D$ and $C:D$ relationships that we wish to highlight here.

As can be seen in Fig. 2, some trends exist in the 1958-79

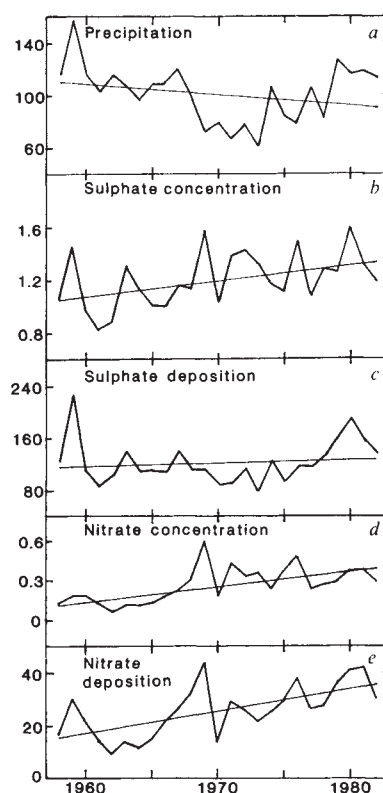


Fig. 2 Eskdalemuir: annual time series of mean monthly precipitation (a), mean monthly NO_3^- concentration (b), mean monthly NO_3^- deposition (c), mean monthly excess SO_4^{2-} concentration (d), and mean monthly excess SO_4^{2-} deposition (e). EACN data for 1958-79 with 1980-82 data from the EMEP data network. Mean monthly values are plotted to compensate for years with some missing data. Rainfall units are mm, concentration units mg l^{-1} and deposition units mg m^{-2} . SO_4^{2-} as S; NO_3^- as N.

EACN data. To minimize the effects of autocorrelation on significance levels the annual data were linearly detrended. Correlations between the above parameters were then reperformed. The results show that the underlying trends are not biasing the relationships. The detrended correlations are also given in the following discussion.

For nitrate, deposition is strongly related to concentration ($r=0.86$, $P<0.001$; detrended, $r=0.80$, $P<0.001$) but not to the amount of precipitation ($r=-0.16$, $P=0.33$; detrended, $r=0.11$, $P=0.63$). For sulphate, the reverse is true with the $P:D$ relationship strong ($r=0.74$, $P<0.001$; detrended, $r=0.77$, $P<0.001$) and a much weaker $C:D$ relationship ($r=0.37$, $P<0.10$; detrended, $r=0.47$, $P<0.05$). Moreover, the sulphate $C:D$ correlation value is increased by one outlier. If that outlier is removed, though analysis of the data does not support its exclusion as an erroneous observation, the $C:D$ correlation becomes nonsignificant.

It appears that while nitrate deposition at Eskdalemuir is 'driven' by the concentration amount, sulphate deposition is driven directly by the amount of associated precipitation. These relationships may arise from the differences in medium in which the oxidation of sulphate and nitrate occurs. Oxidation of NO_x takes place in the gas phase, so amounts of nitrate are likely to be independent of any precipitation amount. Sulphate, in contrast, is likely to be produced in liquid phase reactions⁵, so the amount of sulphate deposited will tend to correlate with precipitation amount. These relationships suggest that any variability in the amount of precipitation alone will produce differential changes in the nitrate and sulphate deposition. Moreover, this would occur irrespective of other factors such as emission rates. An alternative explanation might be systematic changes over

the years in the frequency of precipitation associated with particular airmasses with their own characteristic sulphate/nitrate concentration ratios⁶. It is believed that the geographical location of Eskdalemuir as a receptor, in relation to the pollutant source areas, plays a part in the $P:C:D$ relationships for sulphate and nitrate. Preliminary analysis indicates coherent spatial variations in the relationships over Europe, an aspect that is being studied in more detail for evidence of possible mechanisms.

Time series data for annual precipitation, sulphate concentration and deposition and nitrate concentration and deposition have also been analysed. In the plotting of Fig. 2, data for 1958 to 1979 are from EACN. For 1980 to October 1982 the data are from the European Monitoring and Evaluation Programme (EMEP) network⁷, homogenized to fit with the EACN data.

The rainfall series for Eskdalemuir show no statistically significant linear trend ($r = -0.25$, $P = 0.23$), although there is a period of years with below-average rainfall from 1969 to the late 1970s. Sulphate concentration levels show a positive trend ($r = 0.44$, $P < 0.05$) while the deposition shows no trend ($r = 0.12$, $P = 0.56$). Nitrate concentration levels show a significant trend from 1958 to 1982 ($r = 0.56$, $P < 0.01$); this can also be interpreted as a rising trend to the early 1970s followed by a levelling off¹. The nitrate deposition data show a steady upward trend from 1958 to 1982 ($r = 0.64$, $P < 0.001$). These trends cannot be explained by emission rates alone. UK emissions of SO_2 have declined since the first half of the 1970s, while NO_x emission rose to about 1970 then levelled off^{8,9}. Nevertheless, sulphate concentration levels at Eskdalemuir have risen by 25% during the study period, and nitrate concentration and deposition have doubled. The rise in nitrate deposition is clearly driven by increasing nitrate concentration, while the sulphate deposition is 'suppressed' by the gradual reduction in precipitation totals. A clear example is the period 1967–74 when sulphate concentrations exhibit their major peak (above the trend line) and yet sulphate deposition exhibits a trough, because of lower-than-average precipitation.

This work was funded by the UK Department of Environment and the Commission of the European Community. T. M. L. Wigley made useful comments on an earlier version of this paper.

Received 30 March; accepted 2 July 1987.

1. Rodhe, H. & Rood, M. J. *Nature* **321**, 762–764 (1986).
2. Kallend, A. S. *et al. Atmos. Environ.* **17**, 127–137 (1983).
3. Rodhe, H. & Granat, L. *Atmos. Environ.* **18**, 2627–2639 (1984).
4. Nodop, K., Skjelmoen, J. E. & Schaug, J. *EMEP/CCC-Rep*, 2/85 (Norwegian Institute for Air Research, Lilleström, 1985).
5. Calvert, J. G. *et al. Nature* **317**, 27–35 (1985).
6. Davies, T. D. *et al. Nature* **322**, 359–361 (1986).
7. *The OECD Programme on Long Range Transport of Air Pollutants: Measurements and Findings*. (OECD, Paris, 1977).
8. *Acid Deposition in the United Kingdom* (UK Review Group on Acid Rain, Warren Spring Laboratory, Stevenage, 1983).
9. *UK Photochemical Oxidants* (Review Group Interim Report, Harwell Laboratory, Oxon, 1987).

Improving the accuracy of geomagnetic intensity measurements

D. Walton

Physics Department, McMaster University, 1280 Main Street West, Hamilton, Ontario, Canada L8S 4M1

Many palaeomagnetists believe that, when properly measured, geomagnetic intensities can be obtained with a precision of better than 10% (R. S. Coe, personal communication). But published results do not appear to support this view^{1–5}: the reproducibility of intensities are, not only from laboratory to laboratory^{1,2}, but also within laboratories^{3–5}, in the range 50–100%, whether the Thelliers⁶ method or the Shaw⁷ method is used⁸. As the precision

of laboratory techniques is a few per cent, the most likely source of errors of this magnitude is mineral alteration due to heating during sample processing. Here I describe a technique capable of measuring the effect of alteration, allowing its contribution to be subtracted from the data. This work also shows that these errors can be systematic, confirming a previous finding⁹.

In the Thelliers⁶ double-heating method the sample is heated twice to a succession of increasing temperatures in two different magnetic fields, one of which is usually zero. The ratio of the change in sample moment in the two fields yields the ancient field. This ratio is monitored by plotting the moment remaining after the zero-field step against moment gained. A straight line is equivalent to a constant ratio and is taken to be a sufficient condition for a valid result. But it has been known for some time, both experimentally and theoretically¹⁴ that such a straight line is no guarantee of reliability. The failure of the linearity test is a likely reason that the reproducibility of intensities obtained with Thellier type methods is so poor. Space precludes discussion of the Shaw method.

As there is no technique available to detect alteration that has occurred in antiquity, I shall consider only that which takes place during the laboratory heating necessary for sample processing.

Processes of alteration and magnetization proceed in parallel, and are difficult to separate with the linearity test. But if the sample is heated in a field close to the ancient field so that the

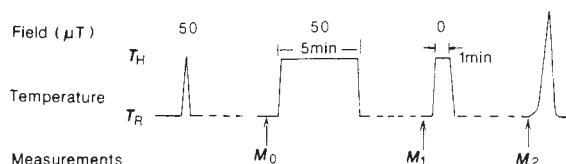


Fig. 1 Schematic representation of the technique employed. All moments were measured at T_R in zero field, but for the first and second steps the sample was cooled in field, and the field was switched off as soon as thermal equilibrium was achieved at T_R .

change in magnetic moment is small¹⁵, any alteration becomes apparent as an anomalous change in moment. The sample moment can be monitored while it remains at a raised temperature, or a series of measurements can be made at separate times, after cooling to some reference temperature. My use¹⁶ of the former procedure was criticized¹⁷ on the basis that it was a new and untested method. I have used the latter in the research reported here. But time and temperature are interchangeable, so the two methods are equivalent.

Tanguy¹⁸ has monitored alteration using a succession of heatings to 700 °C. In my materials this method resulted in too much alteration for a reliable estimate. In contrast to Tanguy, I monitor the change in moment at a series of temperatures in steps of 35 °C or 44 °C to 500 °C. Only two heatings are used because it was found that any further changes were too small to measure. The procedure is a modification of a Thellier method with a short first heating followed by a long second heating in magnetic field, and a third heating in zero field (Fig. 1).

As my samples are recent I have a reliable estimate of the 'ancient field'. The steps are as follows.

(1) Heat (in air) to T_H in a laboratory field of 50 μT and then immediately cool to a lower reference temperature, T_R . Switch field off, and measure the moment, M_0 . To save time, T_R was set as 150 °C.

(2) Heat to T_H once more, hold there for 5 min, and cool again to T_R . Measure the moment, M_1 , as above.

(3) Heat to T_H in zero field, hold there for 1 min and cool to T_R , where the moment, M_2 , is measured.

The moments were measured with a superconducting quantum interference device (SQUID) magnetometer. The samples, which were cylinders 4.5 mm in diameter by 6 mm in length,

were held in a small oven. They were oriented so that their moment was parallel to the axis of the SQUID.

The difference between M_1 and M_0 is a measure of the alteration occurring between the end of the short first step and the long second step. Some alteration may have occurred before the end of the first step, and can be accounted for as follows.

In general the alteration is $f(T)(M_1 - M_0)$, where $f(T)$ is a factor which must be >1 ($f=1$ indicates that no alteration has occurred before the end of the first heating step). For small corrections the temperature dependence of f may be neglected, as here. Alteration is cumulative, so the sum of the corrections at all preceding temperatures must be used. Thus partial thermoremanent magnetization (PTRM) acquired at temperature T_i is $P(T_i) = M_1(T_i) - N(T_i) - f \sum_{k=1}^i [M_1(T_k) - M_0(T_k)]$, where $N(T_i)$ is the natural remanent magnetization (NRM) remaining, obtained from the apparent moment lost, $M_2(T)$, after correcting for alteration; $M_2(T)$ contains the contributions of those grains whose blocking temperature is above some temperature close to T . The fraction of the total number is $\sim M_2(T)/M_1(T)$. Thus $N(T_i) = M_2(T_i) - f \sum_{k=1}^i [M_1(T_k) - M_0(T_k)] M_2(T_k)/M_1(T_k)$.

To estimate f , ten samples from a brick from a London kiln were processed, and f was adjusted until an average field of $50 \mu\text{T}$ was obtained, which is the known field in the kiln after correcting for the different cooling rate in the laboratory^{15,19,20}. The individual fields were obtained from Arai plots of the moment acquired against moment lost at each temperature, after subtracting the signal due to alteration. Figure 2a shows examples. The value of f deduced from this procedure was 2, implying that the alteration during the first heating step was comparable with that between the end of the first and second step. Using this value a number of samples from local bricks were measured and Table 1 shows the results.

The uncorrected field was obtained from an Arai plot of $M_1 - M_2$ against M_2 if at least six points could be used with a standard deviation of $<0.2\%$. The corrected field was obtained by subtracting the signal due to alteration as described above. The substantial reduction in the scatter in the ancient field values is the principal result of this study.

Table 1 Magnetic field (μT) (including a 5% correction for the laboratory cooling rate)

Sample	Uncorrected	Corrected
(a) Bricks		
LBK A	43	51
B	—	48
C	40	46
D	48	50
E	46	49
F	46	48
G	46	47
H	37	46
I	43	47
J	46	50
BK3 A	36	51
B	35	49
DNKO	62	48
SJ38	43	49
DNO	58	50
Average	44.9	48.6
(b) Sherds		
CR74.D2	50	52
CR72.D2	46	53
CR74.G	—	47
CR74.E2	38	47
CR74.A3	36	48
CR74.E3	48	53
CR74.D4	45	53
Average	43.8	50.4

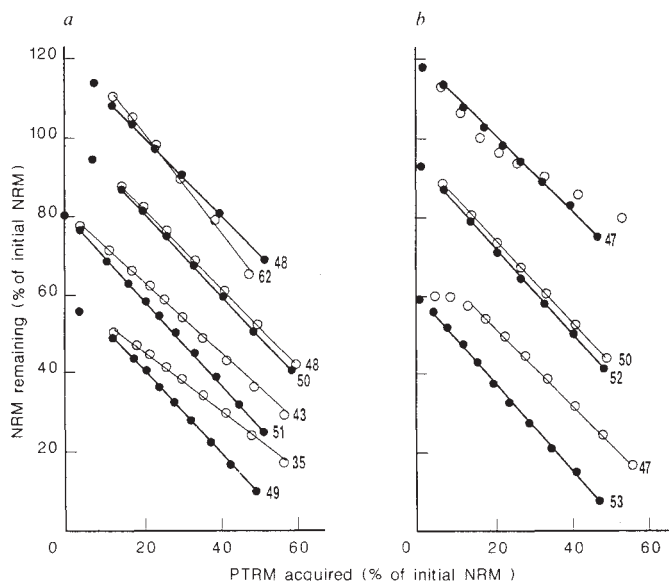


Fig. 2 Representative Arai plots: *a*, bricks (four plots); *b*, sherds (three plots). The open circles are the original data, and the solid circles are the result of subtracting the fictitious moment due to alteration, as described in the text. The numbers are the resulting fields, in μT , after applying a 5% cooling-rate correction. The data have been displaced vertically by an arbitrary amount.

Similar results were obtained for pottery fired in Britain. The pottery was coarse (because fine material was found to be too weak to measure) and was rather difficult to date precisely but was believed to be early 19th century; Table 1 summarizes the results. Figure 2b shows examples of the Arai plots. The average field obtained from the sherds is slightly higher, possibly reflecting an increased dipole moment (observatory data indicate a decrease of about 5% in the last 150 years).

The following conclusions may be drawn from these results.

(1) The reproducibility improves from a range of 35–62 μT , or 60% of the average, to a range of 46–51 μT , 10% of the average for the bricks.

(2) If the uncorrected results are averaged, the field obtained is incorrect (44.9 μT for the bricks and 43.8 μT for the sherds).

(3) A straight line in an Arai plot is no guarantee that alteration has not contaminated the data.

(4) As Fig. 2b shows, departures from linearity in Arai plots can be ascribed to alteration. This of course is not new, but what is new is that samples that would otherwise be rejected can yield excellent results after the signal due to alteration has been subtracted.

In future work smaller samples will be used: this will permit a faster first heating, f will be reduced, and a better estimate for the alteration will be obtained.

Estimating f for ancient material presents a problem: a solution would be to cool a sample in a known field from above the Curie temperature, and then follow the above procedure to estimate f . This would presumably yield a lower limit. Alternatively, if some of the samples reveal no alteration they will yield the correct ancient fields. Then f may be estimated for the others.

Results have been reported²¹ for a series of Egyptian sherds from Tel-el-Amarna, with comparable reduction in scatter. In the method used the zero-field step was introduced between the two in-field steps, and there is a large viscous loss occurring during the second (zero-field) step. The uncertainty involved in accounting for the viscous loss is the main drawback to performing the heating sequence in that order.

This work was supported by the Natural Sciences and Engineering Research Council of Canada. I thank W. Williams

for his assistance, Dr Allison Taylor of the archaeological group of the Cambridgeshire County Council for providing the sherds, and Dr W. H. Owens for suggesting Fig. 1.

Received 17 July 1986; accepted 8 May 1987.

1. Thomas, R. in *Archaeomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 118 (Elsevier, Amsterdam, 1983).
2. Sternberg, R. in *Archaeomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 168 (Elsevier, Amsterdam, 1983).
3. Kovacheva, M. in *Archaeomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 110 (Elsevier, Amsterdam, 1983).
4. Wei, Q. Y. et al. in *Archaeomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 147 (Elsevier, Amsterdam, 1983).
5. Barbetti, M. in *Archaeomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 175 (Elsevier, Amsterdam, 1983).
6. Thellier E. & Thellier, O. *Ann. Géophys.* **15**, 295–296 (1959).
7. Shaw, J. *Geophys. J. R. astr. Soc.* **39**, 133–141 (1974).
8. Kristjansson, L. *Geophys. J. R. astr. Soc.* **80**, 555–559 (1985).
9. Walton, D. *Nature* **315**, 374–375 (1985).
10. Coe, R. S. *J. Geomagn. Geoelectr.* **19**, 157–159 (1967).
11. Coe, R. S. & Gromme, C. S. *J. Geomagn. Geoelectr.* **25**, 415–435 (1973).
12. Prevot, M., Mankinen, E. A., Gromme, C. S. & Lecaillon, A. *J. geophys. Res.* **88**, 2316–2326 (1983).
13. Coe, R. S., Gromme, C. S. & Mankinen, E. A. *J. geophys. Res.* **89**, 1059–1069 (1984).
14. Walton, D. in *Geomagnetism of Baked Clays and Recent Sediments* (eds Creer, K. M., Tucholka, P. & Barton, C. E.) 90–91 (Elsevier, Amsterdam 1983).
15. Walton, D. *Nature* **286**, 245–247 (1980).
16. Walton, D. *Nature* **310**, 740–743 (1984).
17. Aitken, M. J., Allsop, A. L., Bussell, G. D. & Winter, M. *Nature* **314**, 752–753 (1985).
18. Tanguy, J. C. *Earth planet. Sci. Lett.* **27**, 314–320 (1975).
19. Fox, J. M. W. & Aitken, M. J. *Nature* **283**, 462–463 (1980).
20. Williams, W. thesis, Univ. Cambridge (1986).
21. Walton, D. *J. Geomagn. Geoelectr.* **38**, 1349–1352 (1986).

Retreat velocity of the North Atlantic polar front during the last deglaciation determined by ^{14}C accelerator mass spectrometry

Edouard Bard, Maurice Arnold, Pierre Maurice, Josette Duprat*, Jean Moyes* & Jean-Claude Duplessy

Centre des Faibles Radioactivités, Laboratoire Mixte CNRS-CEA, Parc du CNRS, B.P. 1, F-91190 Gif-sur-Yvette, France

* Département de Géologie et Océanographie, Université de Bordeaux 1, Avenue des Facultés, F-33405 Talence, France

The glacial oscillations which dominate the Quaternary global climate are closely related to changes in incoming solar radiation, but the details of palaeoclimatic spectra suggest that other parameters, usually referred to as feedback, have significantly increased the climatic response to extraterrestrial forcing¹. The last deglaciation appears to be one of the more characteristic examples of nonlinearity because its abruptness strongly contrasts with the smooth shape of the associated solar radiation maximum dated at $\sim 11,000$ yr BP. In addition, the presence of the Younger Dryas cold event, concomitant with the solar radiation maximum, could be interpreted as a transitory phenomenon caused by the rapid meltwater influx linked to the first phase of the ice-sheet's disintegration^{2,3}. Here, using accelerator mass spectrometry, we date precisely the movements of the polar front during the whole period of glacial retreat $\sim 15,000$ – $8,000$ yr BP. Our results show that the polar front has moved between 35°N and 55°N in $<1,000$ yr for the earliest retreat $\sim 12,500$ – $12,000$ BP and moved almost 'instantaneously' (that is <400 yr) during the two abrupt climatic changes associated with the Younger Dryas cold event.

Two deep-sea cores have been precisely dated using accelerator mass spectrometry (AMS). The first one was recovered from the Rockall Plateau, offshore Ireland, (CH73-139C; $54^\circ38'\text{N}$, $16^\circ21'\text{W}$) and the second one was raised near the coast of Portugal (SU81-18; $37^\circ46'\text{N}$, $10^\circ11'\text{W}$). The water

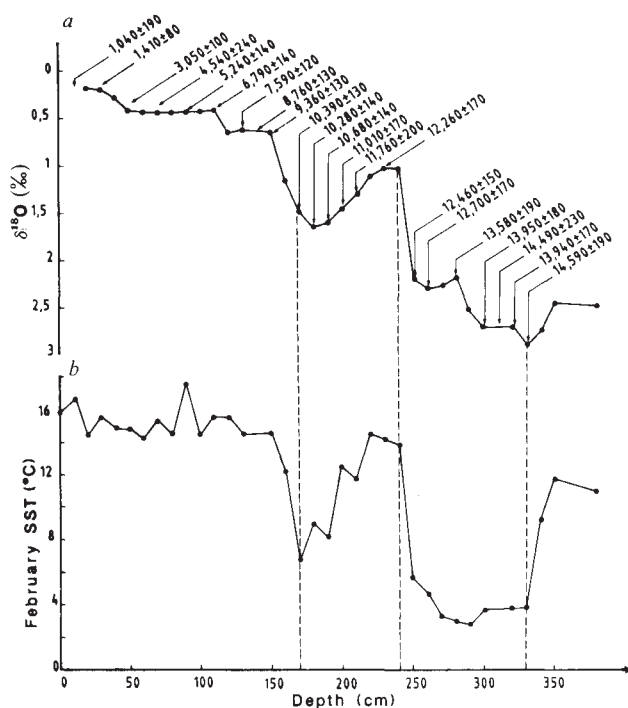


Fig. 1 Climatic records for core SU81-18 (south Portugal). *a*, oxygen isotope and AMS ^{14}C ages obtained on *G. bulloides*; *b*, February sea surface temperature record based on palaeontological transfer function.

depths are respectively 2,209 m and 3,135 m and thus above the average lysocline for these areas. Two kinds of palaeoclimatic signals have been considered in the following approach: the palaeotemperatures based on micropalaeontological transfer functions and the oxygen isotope measurements ($\delta^{18}\text{O}$). Sea surface palaeotemperatures (SST) are estimated by a multivariate statistical analysis of modern and fossil microplanktonic assemblages with a standard deviation approximately equal to $1.5^\circ\text{C}^{4,5}$. The difference between August and February temperatures does not show significant changes through the last deglaciation (for both cores the seasonality is ~ 5 – 7°C). Therefore, for the sake of clarity, we will only mention the February temperature in the following description of our data. The $\delta^{18}\text{O}$ variations measured in planktonic foraminiferal shells are the result of two primary causes: first, the global variations in the isotopic composition of the oceans due to changes in the volume of ice stored over the continents and second, the fractionation between calcium carbonate and water, which strongly depends on the temperature during the shell crystallization. In the case of the last deglaciation the total ice volume effect accounts for $\sim 1.1\%$ ⁶, but the temperature effect is not easy to calculate because the seasonality and living depth during the foraminiferal growth are still poorly known and also because $\delta^{18}\text{O}$ are related to salinity variations. Therefore no attempt has been made here to delineate the different components of the $\delta^{18}\text{O}$ data.

The chronological resolution of the dating of palaeoclimatic signals appears to be the most critical factor as one tries to quantify the velocity of the polar front retreat which followed the last Glacial Maximum (LGM). Fortunately, AMS provides reliable ^{14}C ages with only 1 mg of carbon which enables the direct assignment of precise ages to palaeoclimatic tracer 'carriers' (foraminifera) with an upper limit of $\sim 44,000$ yr BP. This technique cuts out all risk of contamination by older detrital carbonates^{7,8}. To compare our ages with the continental chronologies, all planktonic foraminiferal ^{14}C ages have been corrected for the mean present apparent age of surface water (that is, 400 yr).

Another important problem encountered when attempting to record abrupt climatic changes in deep-sea cores is created by the natural process of bioturbation. In the following development we have used a first order model to evaluate the influence of bioturbation on the palaeoclimatic spectra⁹.

Figure 1 shows the oxygen isotope records measured in the foraminifer *Globigerina bulloides*, the ¹⁴C ages measured in the same species and the palaeotemperature profiles obtained through the transfer function technique for the core SU81-18 (south Portugal). Below 150 cm the sedimentation rate is ~35 cm kyr⁻¹ and after an age of ~9,300 yr BP this rate drops to ~16 cm kyr⁻¹. If we associate the first value with a bioturbation thickness of 10 cm we obtain a theoretical attenuation of ~25% for a signal half-period of 1,000 yr⁹. This signifies that in this particular core the bioturbation smoothing is not a tremendous problem if one considers events longer than 1,000 yr.

Figure 1 shows that Termination IA, defined on the isotope record¹⁰, began ~14,500 yr BP which is consistent with a preliminary study of the deglaciation in core CH73-139C^{9,11}. The SST remained at ~12 °C in winter until just before 14,590 ± 190 yr BP (level 330 cm) and then dropped to values <6 °C in winter as the deglaciation began. Between the levels 250 cm (12,460 ± 150 yr BP) and 230 cm (12,260 ± 170 yr BP), the SST rose to maximum values of 14 °C in winter at a mean rate of ~4 °C per century.

The Younger Dryas cold event is clearly observed in both climatic records and is dated between 11,010 ± 170 yr BP (level 200 cm) and 10,390 ± 130 yr BP (level 170 cm). The polar front readvance can also be evaluated and corresponds approximately to a temperature drop of 0.5–1 °C per century. After the Younger Dryas event, the second phase of deglaciation take place between levels 170 cm (10,390 ± 130 yr BP) and 150 cm (9,360 ± 130 yr BP). The winter SST rise is less abrupt than for Termination IA and corresponds to a mean warming rate of <1 °C per century.

It is difficult to be sure of the past position of the polar front because its actual definition is related to thermal and salinity gradients between different water masses. Considering the sea surface temperature maps drawn for the last glacial maximum^{12,13} it is evident that core SU81-18 is situated in the southern part of the steep thermal gradient which separates the Gulf Stream from the polar water mass. The temperature values of ~12 °C in winter and ~17 °C in summer, calculated from this core before 14,500 yr BP, are in perfect accordance with the CLIMAP reconstruction. In core SU81-18 the beginning of the oxygen isotope decrease is characterized by a concomitant temperature drop of ~7 °C which can surely be attributed to a pulse-like injection of large volumes of ice or melt water into the Atlantic. Following Ruddiman *et al.*¹⁴ we assume that the passage of the polar front across a given core will constitute the dominant thermal event within a general deglaciation warming. In the case of core SU81-18 we can affirm that this oceanographic structure retreated from the latitude of Portugal only 12,500 yr ago.

In the case of core CH73-139C (west Ireland) the interpretation of isotope data is more complicated because the bioturbation disturbance is not negligible. Without taking into account the effect of bioturbation, the raw δ¹⁸O and ¹⁴C data show that the first phase of melting began soon after 15,790 ± 220 yr BP and was complete by ~13,300 yr BP¹¹. The palaeotemperature record (Fig. 2) indicates that winter SST increased between the levels 150–140 cm in the core which correspond to two different AMS ¹⁴C ages: 11,590 ± 150 yr BP measured on the subpolar species *G. bulloides* and 13,950 ± 270 yr BP measured on the polar species *Neogloboquadrina pachyderma* left coiling. Following the interstadial phase, the Younger Dryas cold event has been dated on both foraminiferal species at level 120 cm in depth. The ages are in good agreement and correspond to 10,510 ± 140 yr BP for the subpolar foraminifer and 10,640 ± 160 yr BP for the polar one. The end of the second temperature rise is dated at ~9,200 yr BP on *G. bulloides* and 10,570 ± 210 yr

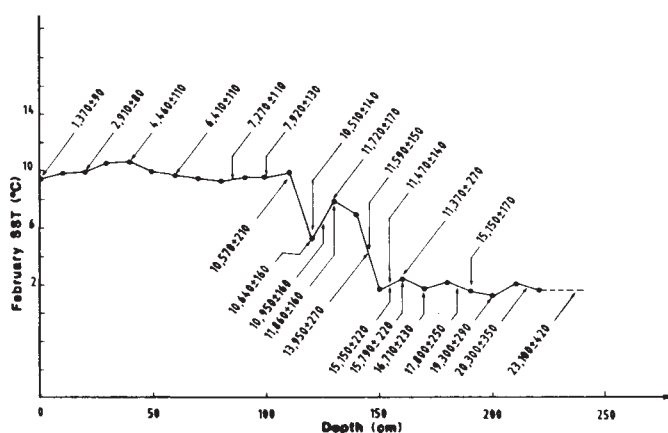


Fig. 2 February SST record and AMS ¹⁴C ages in core CH73-139C (west Ireland). Ages above the curve have been obtained on *G. bulloides* and ages below on *N. pachyderma* left coiling.

BP on *N. pachyderma* (110-cm level). The core top age of 1,370 ± 90 yr BP measured on *G. bulloides* represents a maximum value in that we cannot be sure if sediment was lost during the coring procedure. Nevertheless, this age is consistent with a bioturbation thickness of <16.4 cm (mean sedimentation rate of ~12 cm kyr⁻¹¹¹, 12 cm kyr⁻¹ × 1.37 kyr = 16.4 cm).

In a straightforward interpretation the AMS ¹⁴C ages obtained for these two different foraminiferal species are different but previous studies have shown that bioturbation was responsible for the stratigraphical discrepancies observed in the isotope records: δ¹⁸O and AMS ¹⁴C. A first-order model which takes into account the foraminiferal abundances has been built to simulate and restore the isotope signals⁹. This study has shown that the initiation of Termination IA took place between 14,500 and 15,000 yr BP, the end-part of the deglaciation pause is dated near 11,500 yr BP, the Younger Dryas at 10,700 yr BP and the end of Termination IB at ~9,000 yr BP. This chronology basically fits that of core SU81-18 as well.

Following this approach, we have deconvolved the abundances of 14 planktonic foraminifera with about the same bioturbation parameters which successfully explain the discrepancies between the isotope profiles (equivalent mixing depth of ~10–12.5 cm). The restored foraminiferal data were then introduced in the transfer function programs to calculate an unmixed temperature record. Figure 3 illustrates the restored palaeotemperatures and the micropalaeontological abundances for seven species which represent >95% of the total foraminiferal fauna. To a first approximation it is clear that the polar foraminifer *N. pachyderma* shows abundances in antiphase with those of other species and that all the 'warmer' foraminifera show concomitant abundance maxima situated between the levels 130–140 cm for the unmixed records. All these deconvolved 'pulses' are associated with the restored temperature maximum observed at ~130–140 cm in the core and therefore the timing of this event is the age assigned to the concomitant maxima of foraminiferal abundances.

As shown on Fig. 2 four similar AMS ¹⁴C ages have been measured on *G. bulloides* between the levels of 160 and 130 cm in core CH73-139C. As core SU81-18 does not present an age anomaly for the time period between 11,000 and 13,000 yr BP (Fig. 1), we conclude that a hypothetical change of the atmospheric radiocarbon activity does not explain the constant age in core CH73-139C. Indeed Fig. 3 demonstrates that those levels correspond to the dating of a single 'pulse-like' abundance maximum for *G. bulloides* which has suffered bioturbation between the levels of 130 and 160 cm. The four radiocarbon ages are not statistically different and correspond to a mean value of ~11,540 ± 170 yr BP which appears to be the most

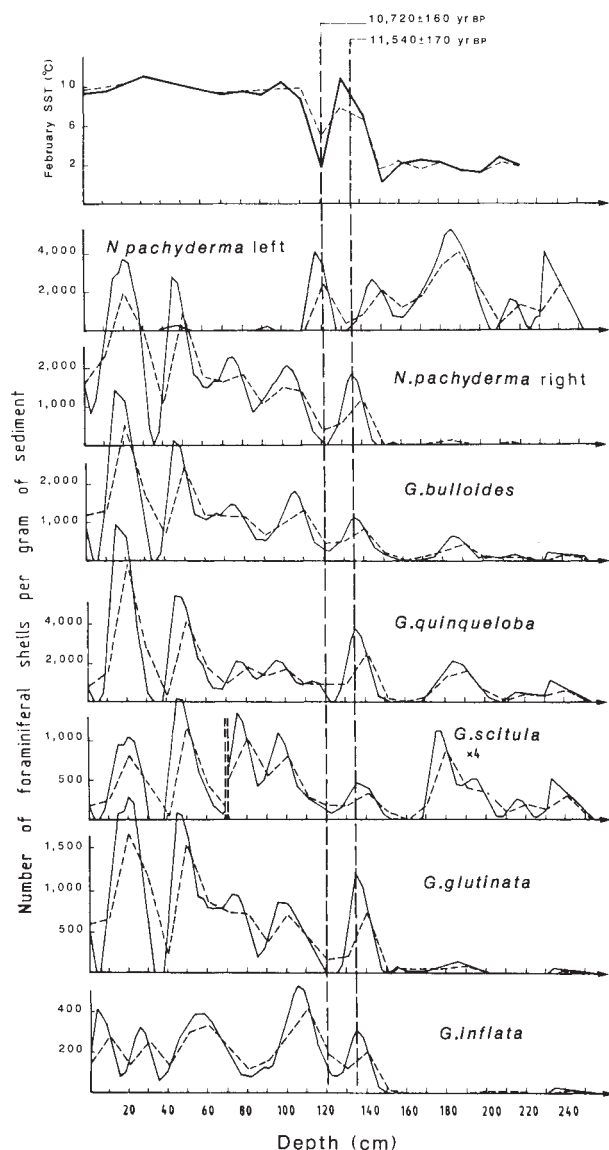


Fig. 3. Core CH73-139C (west Ireland). Upper panel; restored and raw February SST records and two deconvolved ^{14}C ages for core CH73-139C. Dashed line, raw SST record; solid line, restored SST signal. Lower panel, data and deconvolved records for the palaeontological abundances of seven planktonic foraminifera. Dashed line, raw data records; solid line, restored signals. (For *G. scitula* the abundances have been multiplied by a factor of four between 70 and 250 cm in depth).

probable age for the polar front retreat and the invasion of warmer water in the North Atlantic at the latitude of 55°N . Recently Broecker *et al.* (ref. 15 and personal communication) have obtained AMS ^{14}C ages on several foraminiferal species in core V23-81 which is close to core CH73-139C. The timescales of the two cores agree except for the estimated age of the end of the first warming phase. Despite the fact that in core V23-81 the ages obtained at the same levels exhibit a rather large scatter (up to 2,000 yr), the authors conclude that the full interstadial conditions began at $\sim 13,000$ yr BP. Broecker *et al.* do not present the absolute foraminiferal abundances which are the most critical parameters for bioturbation in high-latitude sediments. But, based on the evidence of a volcanic ash layer, the authors believe that the disturbance created by natural mixing is negligible. The eventuality of stratigraphic hiatuses in these high-latitude cores cannot be totally ruled out and we therefore accept that the age of $\sim 11,540$ yr BP for full interstadial conditions

observed in core CH73-139C could be considered as a minimal value.

According to the polar front definition adopted by Ruddiman *et al.*¹⁴, our data show that the North Atlantic polar front began its first retreat $\sim 12,500$ yr ago at the latitude of Portugal and reached the latitude of Ireland at most 1,000 yr later. The distance between the two locations of the cores is close to 2,000 km and therefore the mean velocity of the polar front retreat during the end of Termination IA was $>2\text{ km yr}^{-1}$.

The cooling associated with the Younger Dryas cold event and the warming which followed have been almost 'synchronous' (<400 yr) between 35°N and 55°N , as we do not observe any significant age difference for the temperature variations recorded in both cores. For this time period the errors of one standard deviation (1σ) on the AMS ^{14}C ages are typically ~ 150 – 200 yr, therefore the mean velocity of the polar front during the climatic changes associated with the Younger Dryas may have been $>5\text{ km yr}^{-1}$. This favours the theories that explain the Younger Dryas by instabilities within the climatic system itself, for example, a catastrophic influx of polar ice into the North Atlantic. The high velocity of the advance and the retreat of the cold waters suggest that these temperature changes had affected only the most surficial waters and not the whole hydrological structure of the North Atlantic ocean.

The two AMS ^{14}C records presented here emphasize the extreme rapidity of the climatic changes which occurred during the last deglaciation. Although measurement and stratigraphic uncertainties still exist, this work suggests that it may be possible to demonstrate that the stratigraphic subdivisions of the deglaciation are time-transgressive or synchronous on the European continent. Up till now, the critical limitation for this kind of work was the chronological resolution which is still unsatisfactory in most continental records. Nevertheless, recent studies of pollen time series of deposits situated in the Pyrénées, not far from the Atlantic ocean, indicate that the climate remained rather dry and cold until 13,500–13,000 yr BP^{16,17}. Although the first appearances of hygro-thermophilous species in the pollen-percentage diagrams occurred between 14,500–13,000 yr BP^{16,17}, the real replacement of the open LGM-landscape (herbaceous vegetation) by an extensive woodland (dominated by *Betula* and *Pinus*) can be placed at $\sim 12,500$ – $12,000$ yr BP. These conclusions are particularly evident as one considers the absolute pollen frequencies and the ratio of the arboreal pollens to the total abundance of pollen. Moreover, a recent study of the stratigraphy of a Swiss lake using AMS ^{14}C dating¹⁸, indicated that the age of the early Bölling period was 12,500 yr BP instead of 13,000–13,300 yr BP as expected from the chronozones of Mangerud *et al.*¹⁹ and Welten²⁰. Finally, further north, in the Vosges mountains, Woillard²¹ reported that the retreat of local glaciers was immediately followed by the appearance of Alleröd vegetation in Frère Joseph peat bogs suggesting that the climatic warming at this latitude occurred almost in phase with that observed in core CH73-139C.

The interpretation of the vegetation changes is not obvious because the climatic response of plants is not instantaneous and time lags can also be attributed to the low speed of flora propagation²². To avoid this problem, other authors have studied coleoptera assemblages that are believed to respond almost instantaneously to climatic changes^{23,24}. These studies indicate in Great Britain a rather warm period from $\sim 13,000$ – $12,200$ yr BP followed by a continuous cooling to 11,000 yr BP and further cooling to 10,000 yr BP. Coope^{24,25} reached the conclusion that the Bölling period was as warm as the Preboreal time and that the Alleröd phase was significantly colder than the Bölling, hypotheses which are supported by a few radiocarbon dates measured by β -counting on organic matters. But the possibility of pollution by modern (adsorbed humic acids for example) or dead carbon (hard water effect) has to be qualitatively evaluated for each sample²⁵ and greatly diminishes the validity of such continental chronologies. In contrast, our data indicate that the

high-latitude North Atlantic remained cold until 11,500 yr BP which is in agreement with the observations of Peacock *et al.*²⁶, based on the evidence of benthic foraminifera and molluscan fauna, indicating a brief warm interval which culminated just before 11,000 yr BP on the west coast of Scotland. In addition, several studies indicate that the marine temperatures near Scotland and Norway were significantly colder than today during much of the Bölling/Allerød period²⁷⁻²⁹.

The discussion above indicates that as late as 12,500–11,500 yr BP, the North Atlantic SST remained cold and the landscape on the continent was still open although a limited rise of temperature and humidity can be detected as early as 14,000–13,000 yr BP in the pollen percentages. The possible existence of atmospheric changes in temperature and humidity concomitant with a stagnation of the cold SST which prevailed during the LGM, can be tested by means of general circulation models (GCMs) using appropriate boundary conditions. Unfortunately no simulation of the 11,500–12,500 yr BP situation exists in the literature but the Younger Dryas simulation of Rind *et al.*³⁰ is representative of a climate obtained with cold SST, reduced land-sea ice and relatively high solar radiation. Therefore the comparison between two climate simulations obtained through the same GCM: the 18 K-control simulation (Figs 4 and 6, ref. 31) and the 11KC-0KW simulation (Figs 18c, 19a ref. 30), should qualitatively illustrate the climatic amelioration which might have occurred with a cold Atlantic (18K-CLIMAP SST) in response to changes in other boundary conditions such as incoming solar radiation, land-sea ice and sea-level. The comparison between the surface air temperature simulations: 18 K-control (Fig. 4, ref. 31) and 11KC-0KW (Fig. 18c, ref. 30), indicates a rather small warming over the unglaciated area of western Europe. In contrast, the simulations of precipitation show an increase of ~0.5–1 mm per day on the North Atlantic and western Europe (Fig. 6, ref. 31 and Fig. 19a, ref. 30). These two preliminary conclusions are in accordance with the continental pollen time series discussed previously^{16,17}, which indicate an early colonization by pioneer taxa such as filicales, Umbelliferae and the shrub *Juniperus*.

To achieve a better understanding of the last deglaciation chronology and to assess the existence of a global atmospheric ¹⁴C anomaly, it would be valuable to increase the number and the areal coverage of the AMS ¹⁴C data base obtained on planktonic and benthic foraminifera, to reinvestigate the ¹⁴C dating of continental proxy data such as floral and faunal assemblages observed in lakes and bog cores and also to simulate the climatic conditions which prevailed during the different phases of the last deglaciation. Based on two deep-sea cores, our results suggest that the polar front has moved very quickly during the three major temperature changes of the last deglaciation but we still need more data to build a comprehensive picture of the polar front retreat on the scale of the entire North Atlantic.

We thank J. Antignac, E. Kaltnecker, H. Leclaire and B. LeCoat for help in the analyses; A. Castera, D. de Zertucha and J. Dudouit for their help in developing AMS; E. Kaltnecker for the drawing; and J. Davies, J. Fayolle, G. Jalut, L. D. Labeyrie and D. Peteet for useful discussions and review of the manuscript. This work was supported by CNRS, PNEDC, CEA and the EEC.

Received 26 March; accepted 19 June 1987.

- Ruddiman, W. F. & Duplessy, J. C. *Quat. Res.* **23**, 1–17 (1985).
- Broecker, W. S., Peteet, D. M. & Rind, D. *Nature* **315**, 21–25 (1985).
- Ruddiman, W. F. & McIntyre, A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 145–214 (1981).
- Imbrie, J. & Kipp, N. D. in *The Late Cenozoic Glacial Ages* 71–181 (ed. Turekian, K. K. (Yale University Press, New Haven, 1971)).
- Pujol, C. thesis, Univ. Bordeaux (1980).
- Duplessy, J. C., Labeyrie, L. D. & Shackleton, J. N. *Eos* **66**, 18 (1985).
- Arnold, M., Bard, E., Maurice, P. & Duplessy, J. C. *Nuclear Instr. Meth. Phys. Res.* (in the press).
- Andree, M. *et al. Clim. Dyn.* **1**, 53–62 (1986).
- Bard, E., Arnold, M., Duprat, J., Moyes, J. & Duplessy, J. C. *Clim. Dyn.* **1**, 101–112 (1987).

- Duplessy, J. C., Delibrias, G., Turon, J. L., Pujol, C. & Duprat, J. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **35**, 121–144 (1981).
- Duplessy, J. C. *et al. Nature* **320**, 350–352 (1986).
- CLIMAP Science **191**, 1131–1137 (1976).
- CLIMAP Geol. Soc. Am. Map Chart Ser. **MC-36** (1981).
- Ruddiman, W. F., Sancetta, C. D. & McIntyre, A. *Phil. Trans. R. Soc. B* **280**, 119–142 (1977).
- Broecker *et al. Paleoceanography* (submitted 1987).
- Jalut, G., Delibrias, G., Dagnac, J., Mardonnès, M. & Bouhours, M. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **40**, 321–359 (1982).
- Mardonnès, M. & Jalut, G. *Pollen Spores XXV*, 163–212 (1983).
- Andree, M., Oeschger, H., Siegenthaler, U., Riesen, T. & Moell, M. *Radiocarbon* **28**, 411–416 (1986).
- Mangerud, J., Andersen, S. T., Berglund, B. & Donner, J. J. *Boreas* **3**, 109–126 (1974).
- Welten, M. *Sch. Naturforsch. Gesellschaft.* **95**, 7–104 (1982).
- Woillard, G. *Acta geographica Iovanensia* **14**, 168 pages (1975).
- Watts, W. A. in *The Late-glacial of North-West Europe* 1–21 (eds Lowe, J. J., Gray, J. M. & Robinson, J. E.) (Pergamon, Oxford, 1980).
- Coope, G. R., Brophy, J. A. *Boreas* **1**, 97–142 (1972).
- Coope, G. R., Joachim, M. J. in *The Late-glacial of North-West Europe* 55–68 (eds Lowe, J. J., Gray, J. M. & Robinson, J. E.) (Pergamon, Oxford, 1980).
- Atkinson, T. C., Briffa, K. R. & Coope, G. R. *Nature* **325**, 587–592 (1987).
- Peacock, J. D., Graham, D. K. & Wilkinson, I. P. *Inst. geol. Sci. Rep.* **78/17**, (1978).
- Peacock, J. D. *Boreas* **12**, 73–82 (1983).
- Sejrup, H. P., Høltedahl, H., Norvik, O. & Miljeteig, I. *Boreas* **9**, 203–207 (1980).
- Mangerud, J. *Nor. Geol. Tidsskr.* **57**, 23–54 (1977).
- Rind, D. Peteet, D. *Quat. Res.* **24**, 1–22 (1985).
- Rind, D., Peteet, D., Broecker, W. W., McIntyre, A. & Ruddiman, W. *Clim. Dyn.* **1**, 3–33 (1986).

Cadmium in corals as a tracer of historical upwelling and industrial fallout

G. T. Shen*, E. A. Boyle & D. W. Lea

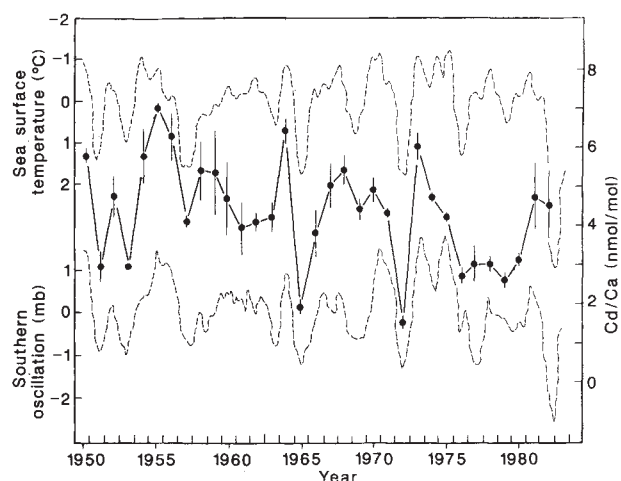
Department of Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 01239, USA

Lattice-bound cadmium in scleractinian corals is a sensitive chemical tracer of historical natural and anthropogenic perturbations to the surface ocean. Cadmium, a biologically-mediated trace constituent of sea water, is distributed analogously to the nutrients, phosphate and nitrate. Under normal circumstances, biological activity depletes surface ocean nutrient and cadmium concentrations. Two corals we have analysed from the Galapagos Islands and North Rock, Bermuda, however, exhibit striking fluctuations in skeletal cadmium within the past century. These variations appear to have occurred in response to two separate phenomena: (1) variable upwelling at Galapagos (El Niño – Southern Oscillation events) and (2) aeolian fluxes of twentieth century, North American industrial cadmium to the western North Atlantic.

Historical Cd/Ca mole ratios in two reef-building corals from San Cristobal Island, Galapagos Islands (*Pavona clavus*) and North Rock, Bermuda (*Diploria strigosa*) are given in Figs 1 and 2a, respectively. Details of trace-metal analysis in aragonitic coral are given elsewhere^{1,2}. Briefly, coral samples were carefully cut to span single annual growth increments (low plus high-density yearly band couplets averaged 11 mm for *P. clavus* and 3 mm for *D. strigosa*). Growth chronologies have been confirmed by seasonal measurements of ¹⁸O in the case of *P. clavus*³ (T. McConnaughey, personal communication) and by stable lead and radiocarbon measurements for *D. strigosa*¹ (E.R.M. Druffel, manuscript in preparation). Cuts were made along narrow high-density bands which are normally accreted during periods of above average water temperature^{4,5}. We cleaned 50–120 mg portions of crushed and sieved coral in a series of oxidizing and reducing media to remove all traces of organic and inorganic surface contamination². The samples were then dissolved in vycor-distilled 2.0 M HNO₃ and lattice-bound trace metals precipitated with APDC (ammonium pyrrolidine dithiocarbamate) using cobalt as a carrier⁶. This procedure isolated metals of interest from the calcium matrix to facilitate analysis by graphite

* Present address: Lamont-Doherty Geological Observatory, Palisades, New York 10964, USA.

Fig. 1 Skeletal Cd/Ca mole ratios in *Pavona clavus*, San Cristobal Island, Galapagos Islands (right axis) compared against historical sea surface temperature anomalies (Puerto Chicama, Peru) and southern oscillation index (Tahiti minus Darwin) (left axes). Annual coral Cd determinations are based on triplicate analyses with resulting standard deviation (σ) errors. Coral was cored at 15-m depth approximately 1 km off Punta Pitt. SST and SOI curves adapted from Rasmussen³¹.



furnace atomic absorption spectrophotometry. Measurement precision ranged from ± 0.1 to ± 0.4 nmol Cd/mole Ca based on triplicate analyses of each annual coral band. Blanks were in the range from 2–6% of measured absorbances in the Galapagos coral and 5–25% for the smaller signals encountered in the Bermuda coral.

We note that the data contained in Figs 1 and 2a define a much smaller range (0.3–7 nmol Cd/mol Ca) than that reported by St John⁷ for four Great Barrier Reef coral genera (n/d–265 nmol Cd/mol Ca). The same is true of the lattice-bound Cd measurements we have produced from the Florida Keys², Mauritius, Eniwetak and Hawaii (unpublished data) where maximum levels do not exceed 3 nmol Cd/mol Ca. Indeed, similar comparisons of our results for other metals (Pb, Mn, Zn)² reveal discrepancies of 1–3 orders of magnitude from past studies^{7–10}. These differences are probably due to inadequate coral cleaning (normally hypochlorite bleaching + distilled water rinse) and trace element analytical techniques employed by previous investigators. Owing to substantial levels of detrital, organic, adsorbed, and handling-emplaced metals in corals, caution must be exercised in interpreting the literature with regard to skeletal distribution coefficients and aragonite lattice compatibility of various trace elements.

In the 33-yr record from Galapagos (Fig. 1) Cd/Ca ratios vary from about 1.5 to 7.0 nmol Cd/mol Ca. A survey of nearshore surface waters at nearby Santiago and Santa Cruz Islands conducted in October 1986 yielded a dissolved Cd concentration range of 15–35 pM. On a San Diego–Panama transect through this same area, Boyle and co-workers observed surface water Cd concentrations varying from ~15 to 75 pM in traversing non-upwelling and upwelling regions¹¹. If one assumes that the Galapagos coral inhabited a zone of variable upwelling, a similar range of surface water Cd concentrations (15–70 pM) is implied by a coral:sea water distribution coefficient (K_D) of 1.0. A similar K_D has been reported for large divalent cations (Sr^{2+} , Ba^{2+} , Ra^{2+}) known to form orthorhombic (aragonitic) carbonates^{8,12,13}. The inferred K_D for Cd in corals is noteworthy in that CdCO_3 is more commonly known as the 6-coordinated rhombohedral calcite structure. Apparently, due to excellent size compatibility with Ca^{2+} (effective ionic radii of 1.12 Å and 1.10 Å for Ca^{2+} and Cd^{2+} , respectively¹⁴) and similarity of electron configuration ($4s^2$ versus $4d^{10}5s^2$), Cd also conforms to an aragonitic structure in nature.

The variable Cd levels recorded by *Pavona clavus* at Galapagos are not random fluctuations. This is evident by comparison with records of sea surface temperature (SST) anomalies and the southern oscillation index (SOI—defined as anomalies in the surface pressure difference between the western and southeastern Pacific), see Fig. 1. Dips in the Cd record reflecting nutrient-depleted conditions appear to correlate with warm water temperatures (recorded at Puerto Chicama, Peru) and a

negative surface pressure anomaly (Tahiti minus Darwin)—both manifestations of El Niño–Southern Oscillation (ENSO) activity. Correspondence is readily apparent for the strong ENSO event of 1972 and the moderate events of 1953 and 1965 (qualitative indices according to Quinn *et al.*¹⁵). Weak events (for instance, 1951, 1969) also appear to follow SST and SOI trends. The record breaking 1982–83 episode is not visible because the coral was cored in 1982 before the full impact of El Niño was manifested. Since the various phases of an El Niño event occur on timescales of months¹⁶, it is likely that much of the detail contained in the SST and SOI records has been filtered out in the annually averaged Cd reconstruction. An example of this is the strong El Niño of 1957–58 which is nearly missed, thus illustrating the need for higher-frequency sampling. Analyses of a sub-sectioned band corresponding to the first half of 1958 gave Cd concentrations similar to those measured for 1965 and 1972. In contrast to the warm nutrient and Cd-depleted El Niño episodes, periods of intense upwelling (for example, 1950, 1955, 1964, 1973) are reflected by unusually cold temperatures and high Cd levels.

The mechanism by which surface nutrient depletion occurs during El Niño stems not from weakened upwelling¹⁷, but from vertical displacement of source waters. As equatorial easterlies weaken during onset of an event¹⁸, the difference in sea-level between the eastern and western equatorial Pacific is diminished¹⁹, causing eastward propagation of Kelvin waves along the equatorial wave guide. These waves act to depress colder, nutrient-rich thermocline waters in the Peru–Galapagos Islands region. Surface ocean warming in the tropical eastern Pacific may occur by advection of 29 °C waters from the western basin²⁰ or by surface heating in the absence of upwelling-induced cooling²¹. Basin-wide temperature perturbations have been recorded in the form of $\delta^{18}\text{O}$ anomalies in corals as far away as Fanning and Canton Island³. As upwelled waters near Galapagos originate from <160-m depth^{22–24}, and upper thermocline gradients for Cd, phosphate, and nitrate are very steep ($\sim 5 \text{ pM m}^{-1}$, $0.016 \text{ } \mu\text{M m}^{-1}$, and $0.3 \text{ } \mu\text{M m}^{-1}$, respectively^{11,25}), a relatively wide range of surface Cd or nutrient concentrations may result, depending on the intensity of vertical restructuring. The range of surface water Cd concentrations inferred from *Pavona clavus* at San Cristobal Island (15–70 pM) suggests a net upwelled source at a depth of only tens of metres (discounting biological effects) in changing from normal to El Niño conditions in this area. Thermocline depression such as observed during the 1982–83 El Niño of 100 m or more²⁶ thus results in catastrophic starvation of the marine ecosystem.

In the Sargasso Sea, at a site 14 km north of Bermuda called North Rock, there is no significant upwelling. This is evident in the very low skeletal Cd levels found in *Diploria strigosa* (Fig. 2a), the highest of which only approach depleted conditions in the eastern Equatorial Pacific. If we assume $K_D = 1.0$

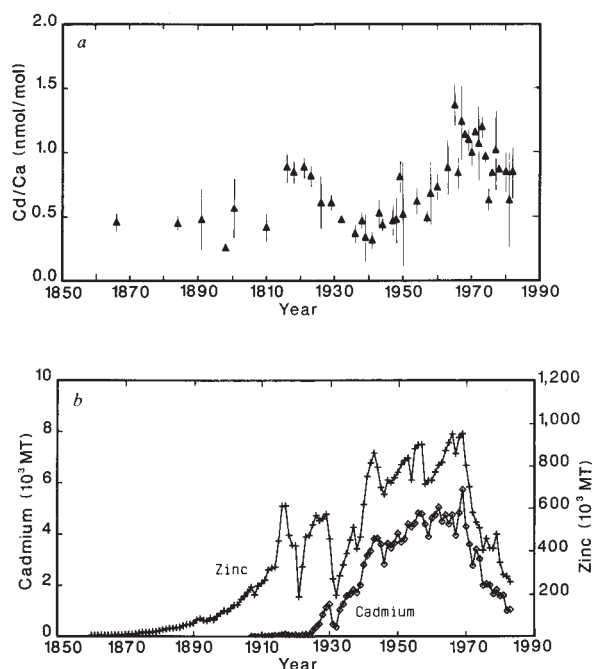


Fig. 2 *a*, Skeletal Cd/Ca mole ratios in *Diploria strigosa*, North Rock, Bermuda. Annual determinations are in most cases triplicate analyses or better with resulting 1σ error bars. Specimen was cored at a depth of 10 m on the seaward side of the reef. *b*, US production records of primary zinc and cadmium: 1850 to the present. (Sources: US Department of Commerce³² and US Department of the Interior—Minerals Yearbook³³).

as in the Galapagos Islands case, surface waters near Bermuda take on values of 3–12 pM. Bruland and Franks reported values close to 2 pM for 1979 Sargasso Sea surface waters²⁷. The coral reconstruction shows anomalously high Cd concentrations to have existed between 1916 and 1932, and 1942 to the present. A two-phase industrial lead rise attributable to long-range aeolian transport of US emissions has been observed in the same coral over the periods 1884–1923 and 1942 to the present¹. Neither the Pb nor Cd perturbations can be attributed to local industrial emissions, as there are none on Bermuda save transportation-related exhausts. Water sampling transects off Bermuda have revealed that island associated dissolved Pb anomalies occur only within two miles offshore¹. The resemblance between the post-1942 coral Cd record and US cadmium metal production (Fig. 2*b*) suggests that the anomalous Cd flux must also be borne by North American westerlies. The older Cd perturbation centred on 1920, however, does not follow the pattern of early industrialization as predictably.

Historically, cadmium has been recovered primarily as a by-product of zinc smelting and to a lesser extent, lead and copper-zinc smelting. Typical Cd/Zn ratios range from 0.3 to 0.7 weight % in various Pb–Zn and Cu–Zn deposits, with particularly high ratios in the Joplin Pb–Zn deposits (0.63%) of Missouri. Initial Cd recovery normally follows either a hydro- or pyrometallurgical procedure, both of which involve Cd collection as flue dust. Comparison of the US cadmium recovery curve with US primary production of Zn in Fig. 2*b* reveals the origin of the unexpectedly high concentrations in the early coral record. Production of Zn rose and fell sharply between 1910 and 1932. Cadmium recovery, which began in 1906, only reached significant levels in 1925. Thus, the coral Cd peak at 1916–23 is a result of a zinc production maximum which occurred at a time when flue dust recovery was negligible. The subsequent Cd decrease in the coral record (1923–32) reflects a combination of declining Zn production and industrial exploitation of Cd-

rich smelting exhausts. Ambient Cd concentrations in surface water as inferred from *D. strigosa*, appear to have regained a 1916-like value only recently (~1970) as Zn production reached its zenith of approximately 10⁶ metric tons per year. A doubling of Zn production between 1916 and 1970, however, implies that about half of the Cd flue dust capable of long-range transport was still escaping to the environment during ore smelting in 1970. This is surprising in view of technological advances in dust recovery²⁸. The explanation lies in decreased contributions of Cd emissions by Zn processing in recent years, relative to other sources (for example, Cu smelting and waste incineration). Nriagu²⁹ estimates that on a global basis, Zn operations accounted for only 40% of total Cd emissions in 1975.

While contemporary industrial perturbation of Pb in the western North Atlantic is of sufficient scale to allow transport modelling³⁰, the industrial Cd signal is prohibitively small to enable a parallel study. Nevertheless, the existence of the observed perturbation for as industrially minor an element as Cd, gives reason to believe that other trace element distributions have been similarly affected. Zinc, nickel, selenium, and copper might be among other likely candidates. Application of Cd in corals as a palaeo-upwelling tracer offers greater potential. As existing records of sea surface temperature, surface pressure and surface marine wind measurements go back only to 1841, an extended coral Cd record would augment our historical perspective of ENSO activity. For example, although the existing record indicates that ENSO events occur aperiodically, an extended record might reveal superimposed longer-period cycles which are presently not discernible. These may in turn implicate forcing mechanisms which are poorly understood. Closer examination of continental records of climate might also expose unforeseen ENSO teleconnections. Further, by performing sub-annual coral measurements of Cd, onset/termination timing patterns over many additional ENSO cycles can be evaluated to improve short-range forecasting models.

We thank E. R. M. Druffel, S. Griffin and T. McConnaughey for providing samples and helpful discussion. This research was supported by the NSF.

Received 29 September 1986; accepted 15 June 1987.

- Shen, G. T. & Boyle, E. A. *Earth planet. Sci. Lett.* **82**, 289–304 (1987).
- Shen, G. T. & Boyle, E. A. *Chem. Geol.* (in the press).
- Druffel, E. R. M. *The Carbon Cycle and Atmospheric CO₂: Natural Variations Archean to Present* (eds Sundquist, E. T. & Broecker, W. S.) 111–122 (American Geophysical Union, Washington DC, 1985).
- Weber, J. N., White, E. W. & Weber, P. H. *Paleobiology* **1**, 137–149 (1975).
- Fairbanks, R. G. & Dodge, R. E. *Geochim. cosmochim. Acta* **43**, 1009–1020 (1979).
- Boyle, E. A. & Edmond, J. M. *Adv. Chem. Ser.* **44–45** (American Chemical Society, Washington, DC, 1975).
- St John, B. E. *Proc. 2nd Int. Coral Reef Symp.* **2**, 461–469 (1974).
- Livingstone, H. D. & Thompson, G. *Limnol. Oceanogr.* **16**, 786–796 (1971).
- Howard, L. S. & Brown, B. E. *A. Rev. mar. Biol.* **22**, 195–210 (1984).
- Milliman, J. D. *Marine Carbonates*, Part 1 (Springer, New York, 1974).
- Boyle, E. A. & Huested, S. *Trace Metals in Seawater* (eds Wong, C. J. et al.) 379–394 (Plenum, New York, 1983).
- Cross, T. S. & Cross, B. W. *J. Sedim. Petr.* **53**, 587–594 (1983).
- Flor, T. H. & Moore, W. S. *Proc. 3rd Int. Coral Reef Symp.* **2**, 555–561 (1976).
- Shannon, R. D. *Acta crystallogr.* **A32**, 751–767 (1976).
- Quinn, W. H., Zopf, D. O., Short, K. S. & Yang, R. T. W. *K. Fish. Bull.* **76**, 663–678 (1978).
- Cane, M. A. *Science* **222**, 1189–1195 (1983).
- Enfield, D. B. *J. geophys. Res.* **86**, 2005–2016 (1981).
- Barnett, T. P. *J. phys. Oceanogr.* **11**, 1043–1058 (1981).
- Wyrtki, K. *J. phys. Oceanogr.* **5**, 572–584 (1975).
- Gill, A. E. & Rasmusson, E. M. *Nature* **306**, 229–234 (1983).
- Newell, R. E. *J. phys. Oceanogr.* **16**, 1338–1342 (1986).
- Quay, P. D., Stuiver, M. & Broecker, W. S. *J. mar. Res.* **41**, 769–792 (1983).
- Fine, R. A., Peterson, W. H. & Ostlund, H. J. *phys. Res.* (submitted).
- Bryden, H. L. & Brady, E. C. *J. phys. Oceanogr.* **15**, 1255–1273 (1985).
- Collier, R. W. *Nature* **309**, 441–444 (1984).
- Barber, R. T. & Chavez, F. P. *Science* **222**, 1203–1210 (1983).
- Bruland, K. W. & Franks, R. P. *Trace Metals in Seawater* (eds Wong, C. J. et al.) 395–412 (Plenum, New York, 1983).
- Wedow, H. *U.S. geol. Surv. Prof. Pap.* **820**, (1973).
- Nriagu, J. O. *Cadmium in the Environment Part I: Ecological Cycling* (ed. Nriagu, J. O.) 35–70 (Wiley, New York, 1980).
- Boyle, E. A., Chapnick, S. D., Shen, G. T. & Bacon, M. P. *J. geophys. Res.* **91**, 8573–8593 (1986).
- Rasmusson, E. M. *Oceanogr.* **27**, 4–12 (1984).
- Bureau of the Census of the US Dept of Commerce *The Statistical History of the United States from Colonial Times to the Present Vols 1965–1984* (Fairfield Publ. Inc., Stamford, Connecticut).
- Minerals Yearbook* (United States Department of the Interior, Washington, DC, 1984).

A long-term increase in the fluid intelligence of English children

Richard Lynn, Susan L. Hampson
& Judith C. Mullineux

University of Ulster, Coleraine, County Londonderry BT52 1SA, UK

In the early decades of the century a number of geneticists and psychologists believed that intelligence in the economically advanced nations was in secular decline^{1,2}. Contrary to this expectation, recent data for a number of countries have shown that intelligence has been increasing at rates far greater than hitherto considered probable. A compilation of this data by Flynn for 14 economically advanced nations has shown that intelligence quotient (IQ) rises have generally been within the range of 2–12 IQ points per decade³. These recent studies raise several questions, among which are precisely what abilities have been increasing over time; and whether existing data are correct in suggesting that the secular rise in Britain is lower than that in other countries. We report here that in English children there has been an increase over the past 50 years of 12.42 IQ points, averaging 2.48 points per decade. The increase has been in fluid intelligence, the mental power that underlies the acquisition of cognitive skills, rather than in crystallized intelligence, which represents the cognitive skills acquired in a particular culture.

To address the question of which cognitive skills have been increasing over time, we adopt the classical British hierarchical model proposed by Spearman and developed by Burt and Cattell. This postulates a broad factor of general intelligence (Spearman's *g*), possibly identifiable with the efficiency of neural processing; two broad factors of the verbal and visuospatial abilities; and some 25–30 narrow primary abilities. Cattell's elaboration of the model splits Spearman's *g* into two general factors identified as fluid intelligence (mental power) and crystallized intelligence (developed cognitive skills)⁴. Fluid intelligence is measured by the culture fair intelligence tests, constructed by Cattell to minimize cultural and educational content and consisting of problems presented in picture and diagram format. The first of these tests was the Cattell Non-Verbal and was standardized for English school children in 1935 (ref. 5). In 1936 the test was administered to all 9–11-year-old children in the city of Leicester. The mean IQ of this sample was 100.49 (ref. 6), which agrees closely with the mean of the 1935 standardization sample and indicates a well standardized test for the measurement of fluid intelligence of English children of the mid-1930s.

In 1985, half a century after the original standardization, we administered the same test to a sample of 1,029 English 9–11-year-old children, drawn from 14 primary schools selected on the advice of headteachers as socially representative and situated in Leeds, Newcastle, Birmingham, Suffolk, Northamptonshire, Cambridgeshire and Sussex. The testing was carried out in the early summer and there were virtually no absentees. Children in special schools for the educationally subnormal were not included in the 1935 sample and were therefore also excluded from the 1985 sample. In 1985 only 1.3% of children attended such schools and the effect of this group on the mean IQ of the total population is negligible⁷.

The results of the 1985 survey showed that in all the schools the children obtained substantially higher IQ scores than the 1935 standardization sample. Working on a standard deviation of 15, the 1985 cohort obtained a mean IQ of 112.42, representing an increase of 2.48 IQ points per decade over the 50 year period. The difference between the means of the 1935 and 1985 samples is statistically highly significant (in Student's *t*-test, $t = 18.2$; $P < 0.001$). There has therefore been an appreciable rise in the IQ of English children in the past 50 years that can be attributed to an increase in fluid intelligence.

In his 14-nation review of secular increases in intelligence, Flynn suggests that the increase in Britain seems to be smaller than that in other economically advanced nations³. This suggestion is based on three sets of data, namely Raven's Progressive Matrices (1938–1979), Cattell's Non-Verbal Test (1936–1949) and the Moray House Test (1932–1947). All three studies produced increases of less than 2 IQ points per decade, whereas virtually all results for other economically advanced nations have shown increases of the order of 2–12 IQ points per decade.

All three of these studies span the years of the Second World War. It may be that this had a depressing effect on intelligence and therefore on the rate of gain. This possibility can be tested, because Cattell administered his test to all 10-year-olds in the city of Leicester in 1949 (ref. 6). If the 1949 mean IQ is used as a base, the increase from 1949 to 1985 was 3.16 IQ points per decade. This figure is appreciably greater than the rise of ~0.75 IQ points per decade for the period 1935–1949. We therefore conclude that the low rates of gain in Britain noted by Flynn are attributable to the inclusion of the Second World War years in the existing studies and that in the post-war period the rate of gain of 3.16 IQ points per decade is within the normal range for other economically advanced nations.

Our results clearly indicate that there has been a substantial increase in Cattell's fluid intelligence in English children over the last half-century. This suggests that the rises in the mean scores on the various tests which have been found in different studies over a wide range of countries are not primarily due to increases in learned cognitive skills resulting from improvements in education and the cultural environment, as proposed by Teasdale and Owen⁸, but in more basic and general mental ability. For further progress in this field it would be useful to investigate and quantify the secular increases that have taken place in other components of the Spearman–Burt–Cattell model.

Received 6 May; accepted 15 July 1987.

1. Fisher, R. A. *The Genetical Theory of Natural Selection* (Dover, London, 1929).
2. Burt, C. L. *Intelligence and Fertility* (Hamilton, London, 1946).
3. Flynn, J. R. *Psychol. Bull.* **101**, 171–191 (1987).
4. Cattell, R. B. *Abilities: Their Structure, Growth and Action* (Houghton Mifflin, Boston, 1971).
5. Cattell, R. B. *A Guide to Mental Testing* (London University Press, 1948).
6. Cattell, R. B. *Eugen. Rev.* **42**, 136–148 (1950).
7. *Statistical Bulletin* (Dept. of Education and Science, London, 1987).
8. Teasdale, T. W. & Owen, D. R. *Nature* **325**, 119–121 (1987).

Adaptive suicidal behaviour in pea aphids

Murdoch K. McAllister & Bernard D. Roitberg

Behavioural Ecology Research Group and Centre for Pest Management, Department of Biological Sciences, Simon Fraser University, Burnaby, British Columbia V5A 1S6, Canada

The host suicide hypothesis¹, which derives from inclusive fitness theory², postulates that parasitized individuals in spatially aggregated populations consisting of close kin may actively enhance their probability of dying. The fitness cost associated with suicide becomes negligible when infection by a parasitoid causes the expected reproduction of the host to approach zero. But the host will benefit from suicide, if by its death (and that of its parasite) the level of subsequent parasitism in its kin is reduced relative to that in non-kin. Although conceptually appealing, host suicide has not yet been clearly demonstrated³. Here we report that pea aphid (*Acyrtosiphon pisum*) parasitized by the braconid wasp *Aphidius ervi*, exhibit apparent suicidal behaviour in response to both aphid alarm pheromone and approaching coccinellid (ladybird beetle) predators. We believe this to be the first convincing evidence in support of the host suicide hypothesis.

Table 1 Mortality probabilities for third instar aphids after dropping from or remaining on plants harbouring a predator

Cause of death	Habitat type	
	Interior (hot dry)	Coastal (cool, moist)
Predation on the plant	<0.25	<0.25
Desiccation on the ground	>0.50	>0.21

Values estimated from refs 11 and 12.

Pea aphids meet several criteria crucial to a test of the host suicide hypothesis. (1) Their viviparous parthenogenetic reproduction and low rate of dispersal lead to aggregations of genetically identical individuals^{4,5}. (2) Aphids parasitized at or before third instar (the preferred stages for parasitism by *A. ervi*) die before producing offspring⁴. (3) Aphid parasites emerge soon (~12 days) after parasitization⁴ and may infect the young of their hosts' siblings. (4) Aphids are susceptible to high rates of predation⁶ when parasitized, before mummification. Thus predation may mediate selection for suicidal behaviour. (5) Aphids display a variety of responses when at risk of predation. These responses are elicited by the visual and tactile stimuli of an approaching predator and by the alarm pheromone released by attacked conspecifics. Each response has different associated survival costs and benefits. Previous work has shown that unparasitized pea aphids can choose that escape response expected to maximize their fitness⁷⁻¹¹. Thus, variation in use of escape tactics occurs in individuals under different environmental conditions and among populations from habitats where payoffs from the different escape tactics vary⁷⁻¹¹.

For the host suicide hypothesis to hold, parasitized pea aphids should alter their escape behaviour such that their probability of death is increased in comparison to unparasitized individuals. Thus, we tested for differences in the ratios of different escape tactics displayed in parasitized and unparasitized aphids under conditions where the associated mortality costs of different escape responses varied.

Pea aphids respond to the alarm pheromone and to approaching predators by backing up—the stylets are withdrawn, the aphid moves backward then remains rigid; by running; or by dropping from the host plant. These responses provide successively greater chances of escape from a nearby attacker but carry successively greater costs in terms of energy spent, lost feeding opportunities and probability of death. The latter is particularly true for aphids in the hot dry climate of interior British Columbia, where the probability of death due to heat stress and desiccation following a drop response may be much higher than the probability of death by predation while on the plant (Table 1). By contrast, ground temperatures in coastal British Columbia are less extreme and the probability of mortality after dropping is much lower (Table 1). Consequently, aphids from the coastal region normally exhibit a higher tendency to drop from plants when faced with predators than those from interior regions^{10,11}. Given the potentially high costs of dropping for interior aphids, evidence for increased dropping by parasitized aphids from those hot dry habitats would support the suicide hypothesis. Further, parasitized aphids from coastal habitats should not increase dropping rates because the risk of death is similar to that faced by remaining on the plant. Parasitized aphids are unlikely to attempt to increase their risk of predation while remaining on their plant because most aphid predators are inefficient: search is random and attack is stimulated by direct chemical and tactile cues¹². Coccinellids in fact often walk right over aphids without eating them.

Pea aphids and *A. ervi* were collected from Chilliwack (coastal, warm moist habitat) and Kamloops (an interior, hot dry habitat) in British Columbia during June and July 1986, respectively. Aphids and wasps from each region were reared separ-

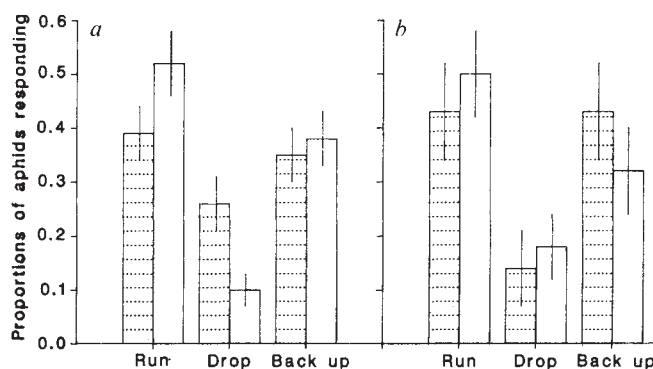


Fig. 1 The escape response of parasitized (lined boxes) and unparasitized (open boxes) pea aphids when exposed to alarm pheromone. Lines show ± 1 s.e.m. range. Note that (s.e.m. = $\sqrt{p(1-p)/N}$, where p = the proportion of aphids responding by the given response type and N = the total number of parasitized or unparasitized aphids over all response types). *a*, Kamloops ('interior', hot dry habitat); *b*, Chilliwack ('coastal', warm moist habitat).

ately on broad bean (*Vicia faba*) and aphids from their matching region, respectively. In the experiments, groups of ten parasitized or unparasitized second instar aphids were placed on the underside of leaves of single broad bean plants with 20 replicates in each case. To avoid bias the backsides of plant pots were labelled either "parasitized" or "control". Groups of interior and coastal aphids were tested alternately on separate days. Test trials were conducted 24 hours following parasitization, at which time most aphids had matured to third instars. During trials temperature and humidity were held constant at $28 \pm 1^\circ\text{C}$ and $33 \pm 3\%$ (hot dry condition), respectively. For the first experiment, third instar aphids obtained from a separate cohort were gently squeezed with microforceps to elicit secretion of cornicle droplets (alarm pheromone) and held 0.5 cm below a group of aphids for 60 seconds. Not all aphids responded to this treatment. Because we could not distinguish whether nonresponding aphids failed to detect pheromone or detected it and chose not to respond, we analysed the data for responders only. Groups of parasitized aphids were dissected afterwards to determine the degree of parasitization and the data for any replicate showing less than 90% parasitism was rejected.

In all experiments, χ^2 tests were used to test for differences between parasitized and unparasitized aphids in the ratios of dropping versus nondropping (running and backing up) individuals.

Aphids from interior regions dropped more frequently when parasitized than when unparasitized (χ^2 with one d.f., $\chi^2_1 = 7.07$, $P < 0.01$; 82 parasitized, 82 unparasitized) (Fig. 1). By contrast, dropping rates for parasitized and unparasitized aphids from coastal regions did not differ ($\chi^2_1 = 0.30$, $P > 0.50$; 28 parasitized, 38 unparasitized). Clearly the increased tendency of 'interior' aphids to drop was not a proximate effect of parasitism itself because coastal aphids did not drop more often after parasitism. These results are consistent with the suicide hypothesis because, as predicted, interior aphids seem to increase the risk of death by desiccation following parasitization whereas coastal aphids, facing alternatives with equal consequence, did not vary their response following parasitism.

To test whether suicidal behaviour in parasitized aphids could be observed under different conditions, we repeated our experiments, this time monitoring the response of parasitized and unparasitized aphids when confronted with an approaching coccinellid predator. Aphids and *A. ervi* were again collected from coastal (Chilliwack) and hot dry interior (Okanagan Falls) areas during September 1986. Coccinellids were maintained at $2-3^\circ\text{C}$. Aphids were treated and placed on plants as in the previous experiment except that this time they were placed on leaves in groups of five. Two days before each trial, coccinellids

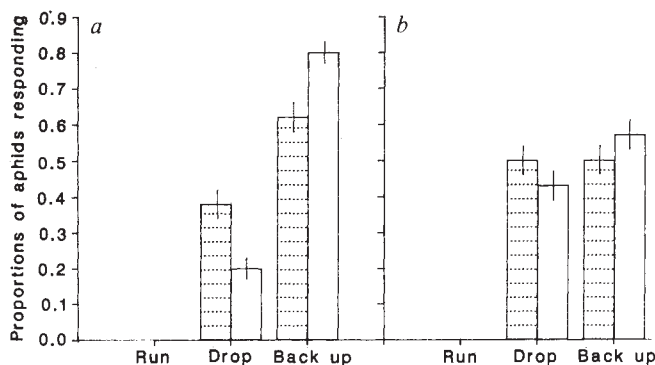


Fig. 2 The escape responses of parasitized (lined boxes) and unparasitized (open boxes) pea aphids when confronted by a coccinellid predator. Lines show ± 1 s.e.m. range. *a*, Okanagan (interior, hot dry habitat); *b*, Chilliwack (coastal, warm moist habitat).

were fed a 50:50 honey/water mixture which was removed 24 h before each experiment to standardize coccinellid hunger. Just before testing, a live coccinellid was gently fastened with melted paraffin to a fine wire attached at one end of a small wooden rod. This tethered coccinellid was manoeuvred along a leaf surface toward individual third instar aphids. Only one coccinellid was used for each group of five aphids. The approach of the coccinellid was repeated as long as aphids appeared undisturbed and still feeding. There were 50 replicate groups for parasitized and unparasitized individuals of each type of habitat. Results were analysed as before.

Once again interior aphids dropped more frequently when parasitized ($\chi^2_1 = 15.2$, $P < 0.0005$; 179 parasitized, 205 unparasitized) (Fig. 2), whereas coastal aphids behaved no differently when parasitized ($\chi^2_1 = 1.73$, $P > 0.20$, 179 parasitized, 191 unparasitized). These results are again consistent with the hypothesis of host suicide. In a habitat where alternative escape responses result in substantial differences in mortality risk, para-

sitized aphids showed the more dangerous alternative. Conversely, in a habitat where alternate responses result in no apparent difference in risk of death, parasitized and unparasitized aphids behaved in the same way. Our results provide support for the hypothesis of host suicide and the notion of suicide as an adaptive behaviour.

Curiously, parasitized aphids do not drop from the plant without mediation by predation. Perhaps incorporation of an already existing stimulus for dropping, presence of the predator, has been the most expedient way for evolution to give rise to the inclusive fitness maximizing behaviours (M. Gross, personal communication). Clearly, dropping behaviour in the absence of predation in unparasitized individuals would be unlikely to persist because it would be detrimental to survival. Consequently, the co-occurrence of this variation with another that would couple it to parasitism may be highly unlikely. Because predator-mediated dropping is an integral part of aphid escape behaviour^{1,3,8-10} and approach by predators frequent^{4,5}, suicidal behaviour may be most easily selected for in the presence of predators.

This study was supported by an NSERC Canada Operating grant (B.D.R.) and an NSERC Canada Undergraduate Scholarship (M.K.M.). We thank B. Lalonde, J. Nelson, M. Reid, L. Dill, M. Gross and R. Ydenberg for improvements to the original manuscript; L. Dill, H. McBrien and M. Mackauer for helpful discussions; and J. Black for technical assistance.

Received 14 April; accepted 13 July 1987.

1. Smith Trail, D. R. *Am. Nat.* **116**, 77-91 (1981).
2. Hamilton, W. D. *J. theor. Biol.* **7**, 1-52 (1964).
3. Stamp, N. E. *Am. Nat.* **118**, 717-725 (1981).
4. Campbell, A. thesis Simon Fraser Univ. (1973).
5. Janzen, D. H. *Am. Nat.* **111**, 586-589 (1977).
6. Hagen, K. S. & van den Bosch, R. A. *Rev. Entomol.* **13**, 325-384 (1968).
7. Brodsky, L. M. & Barlow, C. A. *Can. J. Zool.* **64**, 937-939 (1986).
8. Dill, L. M., Fraser, A. H. & Roitberg, B. D. *J. Anim. Ecol.* (submitted).
9. Muller, F. P. *Ent. exp. Appl.* **34**, 347-348 (1983).
10. Roitberg, B. D. & Myers, J. H. *Can. J. Zool.* **56**, 103-108 (1978).
11. Roitberg, B. D. & Myers, J. H. *Can. Ent.* **111**, 515-519 (1979).
12. Frazer, B. D. & Gilbert, N. J. *Ent. Soc. British Columbia* **73**, 33-56 (1976).

Virus resistance in transgenic plants that express cucumber mosaic virus satellite RNA

Bryan D. Harrison*, Michael A. Mayo* & David C. Baulcombe†

* Department of Virology, Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, UK

† Department of Molecular Genetics, Plant Breeding Institute, Maris Lane, Trumpington, Cambridge CB2 2LQ, UK

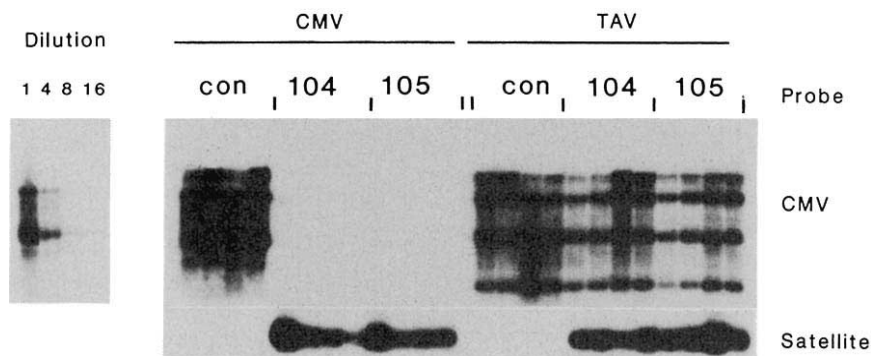
The ability to transform plant genomes with alien genetic material by the *Agrobacterium* Ti plasmid system¹ has made it possible to test whether plant resistance to viruses can be increased by incorporating virus-related nucleotide sequences into the nuclear DNA of the host. We have used this approach² with a satellite RNA of cucumber mosaic virus (CMV) that is replicated only in CMV-infected cells³ and attenuates the symptoms induced by CMV^{4,5}. Tobacco plants transformed with a DNA copy of this satellite RNA produce large amounts of the satellite RNA on infection with a satellite-free inoculum of CMV and, as a result, the satellite becomes transmissible as a component of the virus culture². We report here that CMV replication is greatly decreased and symptom development largely suppressed in these transgenic plants and their sexual progeny, and that tomato aspermy virus (TAV), which is closely related to CMV, likewise induces satellite RNA synthesis

accompanied by symptom suppression but with little or no decrease in TAV genome synthesis. These effects indicate that symptom suppression does not necessarily depend on a decrease in virus replication and that genetically engineered protection by virus satellite nucleic acid is a feasible strategy for enhancing the virus resistance of crop plants.

In many of the experiments described here, the plants were grown from shoot cuttings derived from the original transformants, which contained DNA copies of either 1.3 (construction 104) or 2.3 (construction 105) units of satellite RNA in a head-to-tail arrangement under the transcriptional control of the cauliflower mosaic virus 35S RNA promoter². Plants used in other tests were grown from seed produced by self-fertilization of the original transformants. The seeds were germinated on kanamycin (0.5-1.0 mg ml⁻¹) to select segregants with the T-DNA insertion containing the linked kanamycin resistance gene (neomycin phosphotransferase II) and satellite complementary-DNA constructions.

All the seedlings selected contained small amounts of transcribed satellite RNA and, when infected with satellite-free CMV (strain R76B or strain K), all (over 100 seedlings were tested) produced large amounts of unit-length satellite RNA molecules. Although the amount of transcribed satellite RNA detected in poly(A)⁺ RNA varied several-fold among different non-infected seedlings, similarly large amounts of unit-size satellite RNA were produced in all the seedlings when these were infected with CMV. The ability to produce satellite RNA following infection with CMV was therefore invariably inherited by the seedlings with kanamycin resistance.

Fig. 1 The production of viral RNA and satellite RNA in tobacco plants infected with CMV and TAV. RNA was isolated from systemically infected leaves at 3 weeks after the plants were inoculated. Samples (1 µg RNA) were fractionated by electrophoresis on duplicate formaldehyde-agarose gels and blotted onto nitrocellulose filters. The filters were hybridized with ³²P-labelled probes specific for either CMV RNA or satellite RNA. The probes and the procedures used were as described in ref. 2. The probe for CMV RNA detects all five RNA species of CMV and TAV: RNA-1 and RNA-2 (not resolved in the slowest migrating band) to RNA-5 (fastest migrating band). The plants were grown from seed produced by self-fertilization of the primary transformants and were transformed either with a fragment of the CMV genome as a control (con), with the 104 construction (104) or with the 105 construction (105). Each track contains RNA from a different seedling. As a guide to the magnitude of the inhibition of CMV RNA replication, the left-hand panel shows a sample of RNA from a CMV-infected control plant at loadings of 0.5 (dilution 1), 0.125 (4), 0.063 (8) and 0.031 (16) µg respectively.



In previous tests, the transformed tobacco plants were inoculated with CMV strain R76B, which produces only mild symptoms in control plants, and it was not possible to assess symptom attenuation in the transformed plants. Further tests with a new satellite-free CMV isolate obtained from marrow in Scotland (CMV-K), provided unequivocal evidence of symptom attenuation in the transformed plants. Plants transformed with constructions 104 and 105 developed chlorotic lesions in inoculated leaves, as did control plants. However, in the transformed plants, only the first two or three leaves to be invaded systemically developed mosaic and mottling. Leaves produced subsequently did not show mosaic symptoms and the plants grew almost as well as non-infected ones. In contrast, mosaic symptoms developed in all systemically infected leaves of non-transformed plants or of plants transformed with control DNA, and the plants became obviously stunted (Fig. 2a). These symptom differences persisted for at least 14 weeks after inoculation, the longest period tested, and they were as obvious in plants inoculated with high concentrations of satellite-free CMV (producing more than 200 lesions in an inoculated leaf) as in those inoculated with dilute preparations.

The replication of CMV in the 104-type and 105-type transformants was assayed by three methods. Northern blot analysis with probes for CMV genomic RNA and for satellite RNA was used to assess the effect of the satellite RNA on the production of viral genomic RNA. Production of viral coat protein was determined by serology, using the plate-trapped antigen form⁶ of enzyme-linked immunosorbent assay (ELISA). Finally, tests of infectivity on *Chenopodium amaranticolor* were used to deter-

mine the titre of biologically active virus in the infected plants. *C. amaranticolor* is a local-lesion host of CMV in which the number of lesions that develop in inoculated leaves indicates the titre of virus in the inoculum. The three tests produced parallel results showing, as with the production of symptoms, that the effects of satellite RNA were manifest only in the systemically infected leaves. These effects include decreases of at least 80% (usually about 95%) in the concentrations of CMV genomic RNA and RNA-4 (Fig. 1) and of CMV coat protein (ELISA results, not shown). In addition, the infectivity of leaf extracts from the 104-type transformed plants was much decreased compared with extracts of control plants (Table 1). These results indicate therefore that the suppression of CMV symptoms in the plants transformed with satellite nucleic acid is accompanied by, and may depend (at least in part) on the inhibition of CMV replication.

To test the virus specificity of these effects, 104-type transformed and control plants were inoculated with viruses from a range of taxonomic groups. These included brome mosaic bromovirus and alfalfa mosaic virus, which resemble CMV in having a tripartite RNA genome; tobacco ringspot nepovirus and tomato bushy stunt tombusvirus, some cultures of both of which have a satellite RNA of their own; and peanut stunt virus and TAV which, like CMV, are cucumoviruses. Indeed, TAV is so closely related to CMV that some combinations of genome parts of the two viruses are viable⁷, and TAV can support the replication of at least one CMV satellite RNA⁴. Of these viruses only TAV induced the synthesis of satellite RNA in the 104-type transformed plants. In fact, following infection by TAV there was as much unit-length satellite RNA as in equivalent plants infected with CMV (Fig. 1). TAV was also the only other virus with which modified symptoms were induced in the plants transformed to express satellite RNA. As with CMV, the first two to three leaves to be infected systemically with TAV developed symptoms similar to those in control plants. Leaves produced subsequently developed only slight, but discernible, symptoms in contrast to the obvious yellow mosaic shown by control plants (Fig. 2b). Moreover, the transformed plants grew much more strongly than control TAV-infected ones and almost as well as non-infected plants. Furthermore, the satellite RNA became transmissible as a component of the TAV culture, which then induced milder symptoms in tobacco and *Nicotiana glauca* (Fig. 2c). However, despite these effects on symptoms, neither the accumulation of TAV genomic RNA (Fig. 1) nor the infectivity of leaf extracts (Table 1) was noticeably decreased in the 104-type transformed plants. This result suggests that satellite RNA neither inhibits TAV RNA replication, nor interferes with the production of intact virus particles. In this example, therefore, the ability of satellite RNA to attenuate symptoms is separated from the gross effects on virus replication

Table 1 Infectivity of sap from systemically-infected leaves of transformed tobacco plants

Sap dilution	Control A	Plant type Control B	104-transformed
Cucumber mosaic virus			
1/10	—	—	37
1/50	197	224	9
1/250	29	49	2
Tomato aspermy virus			
1/50	401	580	616
1/250	39	66	47

Sap from systemically infected leaves, sampled 6 weeks after the plants were inoculated, was diluted in 0.03 M phosphate buffer, pH 7.4, and inoculated to corundum-dusted leaves of *Chenopodium amaranticolor* (six leaves per inoculum). Values are mean no. of lesions per leaf. Control A plants were transformed with part of the tobacco rattle virus genome, control B plants with the CMV coat protein gene. CMV coat protein was not detected, however, in preliminary tests on non-infected control B plants.

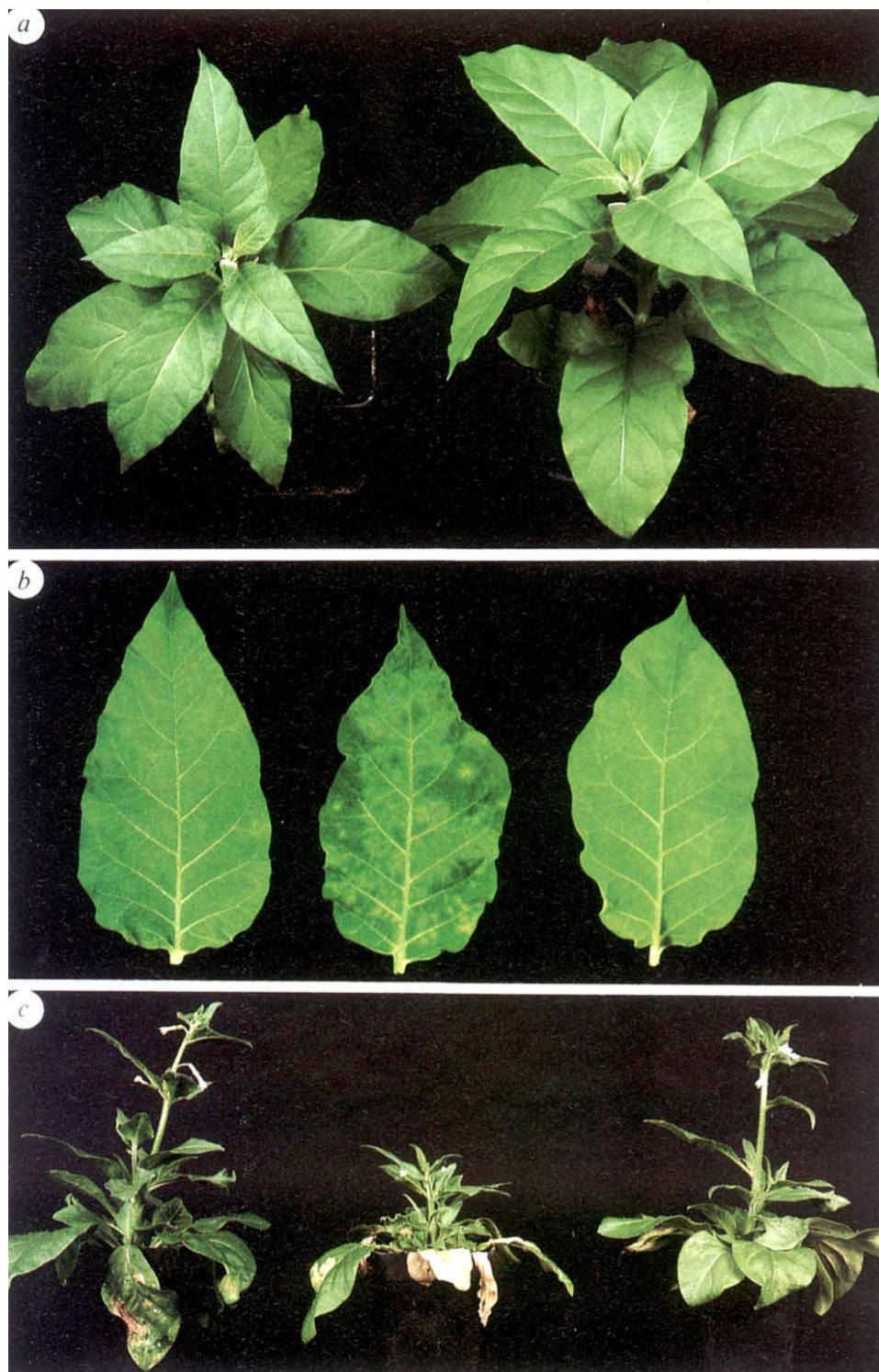


Fig. 2 Effects of transformation with satellite nucleic acid on symptoms of virus infection. *a*, Relative effect of CMV strain K on (left) control tobacco plant transformed with a nucleotide sequence derived from the genome of tobacco rattle virus, and (right) 104-type transformed plant, which shows no mosaic and is growing more strongly. *b*, Effects of TAV on systemically infected tobacco leaves from (left) 104-type transformed plant, (right) 105-type transformed plant and (centre) control plant, as in *a* above. The leaves from the satellite-transformed plants lack the typical yellow mosaic induced by TAV (centre). *c*, Effect on *Nicotiana clelandii* plants of TAV isolates subcultured from (left) 105-type transformed tobacco, (centre) control plant, as in *a* above, and (right) 104-type transformed plant. The TAV culture has become less virulent as the result of acquiring the CMV satellite RNA during propagation in the satellite-transformed plants.

that were found with CMV. It remains possible, however, that satellite RNA has a slight effect on the early stages of TAV RNA replication in cells. If the production of symptoms is dependent not on the final amount of viral RNA accumulation (Fig. 1) but on its timing in relation to cell development, then a slight delay in virus replication could explain the modification of TAV symptom induction by satellite RNA.

Other explanations for the symptom attenuating ability of satellite RNA might involve the direct interactions observed between satellite RNA and viral RNA⁸ or the possible production of small satellite-encoded proteins which interfere with symptom induction by the virus⁹.

As a strategy for preventing the considerable crop losses caused by CMV and TAV^{10,11}, transformation with satellite

nucleic acid has some advantages and some disadvantages. It has the advantage, over the deliberate dissemination of satellite-containing CMV isolates practised in China¹², that yield losses will occur only after natural infection with the virus and these will then be less than those for non-transformed plants. Moreover, satellite RNA will only become transmissible to other plants after the transformed plants have become infected with CMV or TAV, and the virus will then be modified to a less virulent form. Possible disadvantages of transformation with satellite nucleic acid are that the satellite, although benign in the source plant, might be virulent in another species or might mutate to a form which is virulent in the source species. Indeed the same satellite RNA can have very different effects in different CMV-infected hosts⁵, and benign and virulent satellites can

differ in only a few nucleotides¹³. Future work will therefore attempt to disarm the satellite such that it is no longer capable of transmission from the transformed plant.

An alternative strategy for transformation of plants to enhance their virus resistance has been described in which the viral coat protein is produced in the transformed plants. This approach, which is an attempt to simulate natural cross-protection between virus strains^{14,15}, seems applicable to a wide range of viruses. However, it differs from the approach we have adopted both in its concept and its effects. Most importantly, the coat protein strategy results primarily in symptom retardation and is sensitive both to the strength of the viral inoculum (in the range from 0.4 to 2.0 $\mu\text{g ml}^{-1}$ for tobacco mosaic virus) and to the amount of coat protein expressed¹⁴. This, presumably, is because the effect of the expressed coat protein is overcome when the viral RNA accumulates above a certain concentration. Satellite RNA, in contrast, is a molecule that is amplifiable by the incoming virus. Thus, although the system does not give complete inhibition of virus replication, and indeed requires the virus to replicate for the satellite to be replicated, the protection is permanent, independent of the strength of the viral inoculum, and insensitive to the concentration of satellite transcript in the transformed

plants. Further work is required to explore whether this strategy of transformation with satellite can be applied to other plant viruses.

We thank Joanne Clark, Anne Jolly, Elizabeth Murrant and Graham Saunders for technical help. The work was supported in part by the Rockefeller Foundation and the Agricultural Genetics Company Ltd.

Received 27 May; accepted 29 June 1987.

1. Bevan, M. W. & Chilton, M.-D. *A. Rev. Genet.* **16**, 357-384 (1982).
2. Baulcombe, D. C., Saunders, G. R., Bevan, M. W., Mayo, M. A. & Harrison, B. D. *Nature* **321**, 446-449 (1986).
3. Linthorst, H. J. M. & Kaper, J. M. *J. gen. Virol.* **66**, 1839-1842 (1985).
4. Mossop, D. W. & Francki, R. I. B. *Virology* **95**, 395-404 (1979).
5. Waterworth, H. E., Kaper, J. M. & Tournant, M. E. *Science* **204**, 845-847 (1979).
6. Mowat, W. P. & Dawson, S. J. *virol. Meth.* **15**, 233-247 (1987).
7. Habili, N. & Francki, R. I. B. *Virology* **61**, 443-449 (1974).
8. Rezaian, M. A. & Symons, R. H. *Nucleic Acids Res.* **14**, 3229-3239 (1986).
9. Owens, R. A. & Kaper, J. M. *Virology* **80**, 196-203 (1977).
10. Quiot, J. B., Douine, L. & Sélassié, K. G. *Ann. Phytopathol.* **11**, 283-290 (1979).
11. Kaper, J. M. & Waterworth, H. E. *Handbook of Plant Virus Infections and Comparative Diagnosis* (ed. Kurstak, E.) 257-332 (Elsevier, Amsterdam, 1981).
12. Tien, P. *et al. Ann. appl. Biol.* (in the press).
13. Gordon, K. H. J. & Symons, R. H. *Nucleic Acids Res.* **11**, 947-960 (1983).
14. Powell Abel, P. *et al. Science* **232**, 738-743 (1986).
15. Tumer, N. E. *et al. EMBO J.* **6**, 1181-1188 (1987).

Construction of a plant disease resistance gene from the satellite RNA of tobacco ringspot virus

Wayne L. Gerlach, Danny Llewellyn & Jim Haseloff

CSIRO Division of Plant Industry, GPO Box 1600, Canberra, ACT 2601 Australia

Tobacco ringspot virus (TobRV) is the type member of the nepoviruses¹. It consists of 28-nm isometric particles which contain one or the other of the two single-strand genomic RNAs of 4.8 and 7.2 kilobases (kb) (refs 2 and 3). TobRV infects a wide range of dicotyledonous plants and is the causative agent of the budblight disease of soybean. A small RNA which can replicate to high levels and be encapsidated by TobRV in infected plants has been found during serial passages of virus isolates^{4,5}. It is not required for virus propagation and has no detectable sequence homology with the virus genomic RNAs. It is therefore termed the satellite RNA of tobacco ringspot virus (STobRV). It can be considered a parasite of the virus and it ameliorates disease symptoms when present during virus infection of plants. We report here the expression of forms of the STobRV sequence in transgenic tobacco plants. Plants which express full-length STobRV or its complementary sequence as RNA transcripts show phenotypic resistance when infected with TobRV. This is correlated with the amplification of satellite RNA to high levels during virus infection of plants.

The complete nucleotide sequence of the budblight isolate of STobRV is known⁶, and does not appear to encode functional polypeptide products. As is conventional, we will refer to the polarity of the packaged RNA strand as (+), and that of the complementary strand as (-). STobRV replicates through (-)strand intermediates⁷, in an apparently similar way to other related circular satellite RNAs (refs 8 and 9), and 'rolling circle' mechanisms have been proposed to account for the concatameric (+) and (-)strand transcripts found in infected tissues⁷. Both (+) and (-)strand STobRV concatamers undergo self-catalysed ribonucleolytic cleavage at unique sites *in vitro*¹⁰⁻¹², and may do so *in vivo*, to produce unit-length STobRV sequences. Inoculation of TobRV infected plants with autolytically-cleaved *in vitro* transcripts of (+)STobRV resulted in the propagation of STobRV and amelioration of virus induced symptoms¹³. In contrast, (-)STobRV transcripts showed no biological activity.

We have now prepared synthetic plant gene constructions, designated SH+, ST+ and ST- according to the insert, using cloned STobRV complementary DNA segments (Fig. 1). SH+ contains an internal *Hae*III fragment from cloned STobRV cDNA (ref. 13) in an orientation which allows transcription to produce (+) strand RNA, whereas ST+ and ST- contain inserts which are trimers of the cloned permuted monomer cDNA of STobRV in orientations which allow transcription of (+) and (-)STobRV sequences, respectively. These inserts are fused to a constitutive promoter element from the cauliflower mosaic virus (CMV) 35S promoter and 3' terminal sequences from the nopaline synthase gene of *Agrobacterium tumefaciens* (Fig. 1). They were introduced into *Nicotiana tabacum* var. Samsun using *A. tumefaciens* T-DNA transformation and selection with a kanamycin resistance (*kan*^r) marker.

Transgenic plants were regenerated and tested for expression of STobRV transcripts by RNA blot analysis of leaf RNA. Plants producing detectable levels of transcript were chosen for further study. Autoradiographs prepared from RNA blots of transgenic seedling progeny of these regenerated plants (Fig. 2) show hybridization signals similar to those seen in tests of RNA from the original regenerants. The amounts of detectable STobRV transcripts in plants containing SH+ and ST+ STobRV gene constructions were higher than those containing the ST- construction. RNAs detected in the plant containing the SH+ gene construction were of ~1,500 bases, as expected for polyadenylated transcripts from this gene. The major transcript from the ST+ gene was STobRV RNA of monomeric length (359 bases), consistent with autolytic cleavage of the gene transcript. Minor bands in the autoradiograph are made up of products of partial autolytic cleavage. None of the regenerated plants containing the ST- construction produced RNA transcripts at comparable levels. A transformed plant which produced low levels of transcript was chosen for further study. The major transcript found in this plant was monomer length (359 bases), as expected from autolytic cleavage of concatameric (-)strand transcripts. Additional minor bands resulting from partial cleavage products and nonspecific binding to the excess of plant rRNA were also evident on the autoradiograph. These results provide the first proof that the autolytic cleavage of STobRV (+) and (-)strands seen *in vitro* can also occur *in vivo*. Further, in the transgenic plants there is no evidence of RNA-dependent RNA transcription in the absence of virus because only the transcript strands are detected using single-strand specific hybridization probes in RNA blots (Fig. 2).

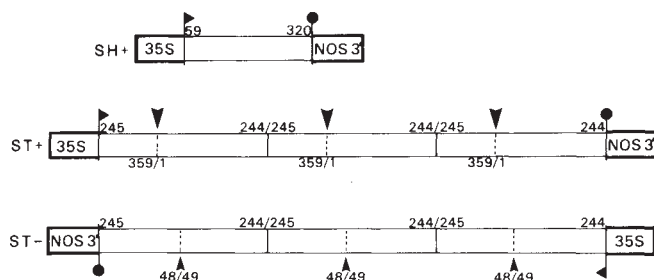


Fig. 1 Schematic representation of synthetic genes in transgenic tobacco plants (not to scale). For all genes, the CaMV 35S promoter and 3' terminal transcription termination and polyadenylation signals of the nopaline synthase gene¹⁸ (NOS 3') were used for expression of STobRV sequences. Gene transcripts consist of RNAs containing a short leader sequence (4 bases) and the inserted STobRV sequence, followed by the 3' transcribed and polyadenylated region of the NOS gene (~1,200 bases). Nucleotides of the insert are numbered from 1 to 359 according to the sequence of the (+)strand of the satellite⁷, and the gene constructions designated according to their cDNA insert as follows: SH+ (top) contains a *Hae*III fragment of STobRV cDNA, including nucleotides 57–319 of the sequence, and oriented to produce (+)strand transcripts. It lacks sequences required for autolytic cleavage *in vitro*. ST+ (middle) contains three direct repeats of a *Sau*3AI fragment of STobRV cDNA (refs 6 and 13), each repeat corresponding to a permuted full-length copy of STobRV, and oriented to produce (+)strand transcripts. The duplication of autolytic cleavage sites (between nucleotides 359 and 1) allows production of correctly terminated (+)strand STobRV RNA monomers. ST- (bottom) contains the trimer sequence described for ST+ oriented in the opposite direction. Duplication of (–)strand autolytic cleavage sites (between residues 48 and 49) allows production of (–)strand STobRV RNA monomers. Potential autolytic cleavage sites for both (+) and (–)strand RNAs are indicated by arrows. ▶, ●, respectively, 5' and 3' extents of the transcribed STobRV sequences.

Methods. The 35S expression vector p35SN was constructed by replacement of the NOS promoter of the expression vector pLGV2383 (ref. 18) with a *Bal*31 deleted CaMV 35S promoter extending from nucleotides –418 to –1 (ref. 19). Unique sites for *Bam*HI, *Sma*I and *Eco*RI are located downstream of the 35S transcription start and allow for insertion of sequences to be constitutively expressed in plant cells. Multimers of the *Sau*3AI insert from the cloned permuted monomer cDNA of STobRV (refs 6 and 13) were cloned into the *Bam*HI site of p35SN to produce permuted trimer constructions of both orientations. The *Hae*III fragment was also cloned into the *Bam*HI site using *Bam*HI linkers. Constructions were verified by restriction analysis and sequencing of the insert junctions. Each of the constructions was co-integrated between the T-DNA borders of the binary vector pGA470 (ref. 20) at unique *Hind*III sites and transferred to the *A. tumefaciens* strain LBA4404 (ref. 21) by triparental mating²². Constructions in *Agrobacterium* were verified by restriction and DNA blotting analysis. Leaf disks of *Nicotiana tabacum* var. Samsun line 5 (ref. 23) were infected with bacteria containing each of the constructions and transgenic plants recovered by selection on shoot regeneration medium containing 100 μ g m^{–1} kanamycin²⁴. Individual transformants with roots were analysed for expression of satellite RNA sequence transcripts by RNA blot analysis and then transferred to soil in a biosafety containment glasshouse. Plants were self-fertilized and R₁ seed collected for progeny analysis and virus infection testing of R₁ seedlings.

Seeds were collected from the regenerated plants. Stable incorporation of the transformed DNA was checked by 3:1 segregation of the linked kan^r phenotype in plants germinated from the seeds. To determine the effects of virus infection on the transgenic plants, twelve seedlings from each transformed genotype and twelve control (nontransgenic) seedlings were inoculated with TobRV virus four weeks after germination. Virus infection was established in all plants as judged by lesion development and by nucleic acid hybridization of leaf extracts using virus genome cDNA probes (data not shown).

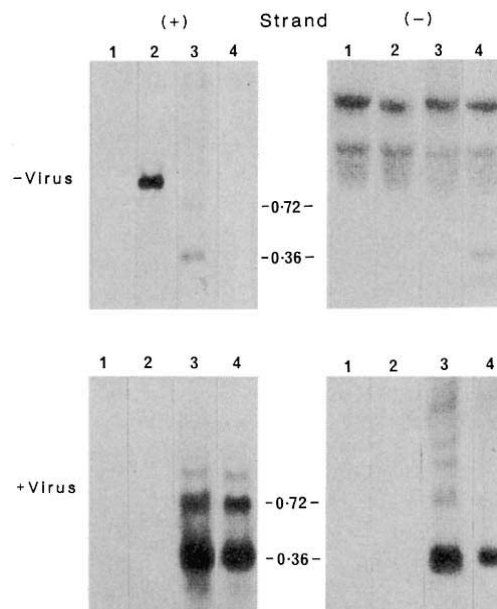


Fig. 2 RNA blot hybridizations showing STobRV RNAs in transgenic seedlings from regenerated plants. Samples are from infected leaves three weeks after inoculation with virus (+ Virus), compared with uninfected controls (– Virus). Single-strand probes were used to detect plus (+) and minus (–) strand sequences of STobRV. Results for representative plants of each genotype are shown. Lane 1, non-transgenic *N. tabacum* var. Samsun; lane 2, SH+ transgenic plant; lane 3, ST+ transgenic plant; lane 4, ST- transgenic plant. Mobilities of STobRV monomer (0.36 kilobases (kb)) and dimer (0.72 kb) size markers are shown.

Methods. RNA was extracted from leaf tissue (3 g) of representative plants. The leaf tissue was homogenized in a mixture of 4 ml buffer-saturated phenol, 8 ml chloroform, 4 ml 0.1 M NaCl, 0.1 M Tris Cl pH 9.5, 10 mM EDTA, 10% SDS, 140 mM 2-mercaptoethanol in a Sorvall Omnimixer (30 s, maximum speed). We added 400 μ l 3 M sodium acetate pH 5.5 and the sample was centrifuged to separate phases and pellet particulate matter. The aqueous phase was recovered and nucleic acids were precipitated with 3 vols ethanol. Pellets were resuspended in 0.8 ml H₂O and RNA samples (40 μ g) were electrophoresed in 6% formaldehyde, 1.5% agarose gels²⁵. Following two 30-min washes of the gel in distilled water, the nucleic acids were blotted overnight to Schleicher and Schuell BA45 nitrocellulose using 20 $\times$ SSC transfer solution. After transfer the filter was baked under vacuum at 80 °C for 1.5 h. The hybridization method was that of Boyer²⁶ with modifications, using unlabelled single-strand M13 probes containing trimer (+) or (–) STobRV cDNA inserts for strand specific hybridization and radio-labelled M13 replicative form (RF) (10⁸ c.p.m. μ g^{–1}) as radioactive probe. X-ray film (Fuji RX) was exposed to the filters as follows: top exposures (– Virus): 1 day (plus strand) and 5 days (minus strand) exposure at –80 °C with intensifying screen; lower exposures (+ Virus): 10 min (plus strand) and 120 min (minus strand) at room temperature. The bands of high relative molecular mass seen for uninfected plant extracts hybridized with (–)strand probe are thought to result from non-specific trapping of probe on the filter by the vast excess of plant ribosomal RNA sequences. The bands comigrate precisely with those of the stained rRNAs and are also evident on long exposure of filters hybridized with (+)strand probe.

STobRV (+) and (–) RNAs were assayed in virus-infected plants using single-strand cDNA hybridization probes (Fig. 2). After infection with TobRV, untransformed tobacco plants and the SH+ transgenic plants showed no alteration in levels of satellite RNA sequences, compared to uninfected plants. However, in plants containing the ST+ and ST- gene constructions there was amplification of satellite RNA sequences to high levels. This is consistent with replication of the STobRV sequence during the infection of the plants with virus.

A striking feature of the virus infections was the altered

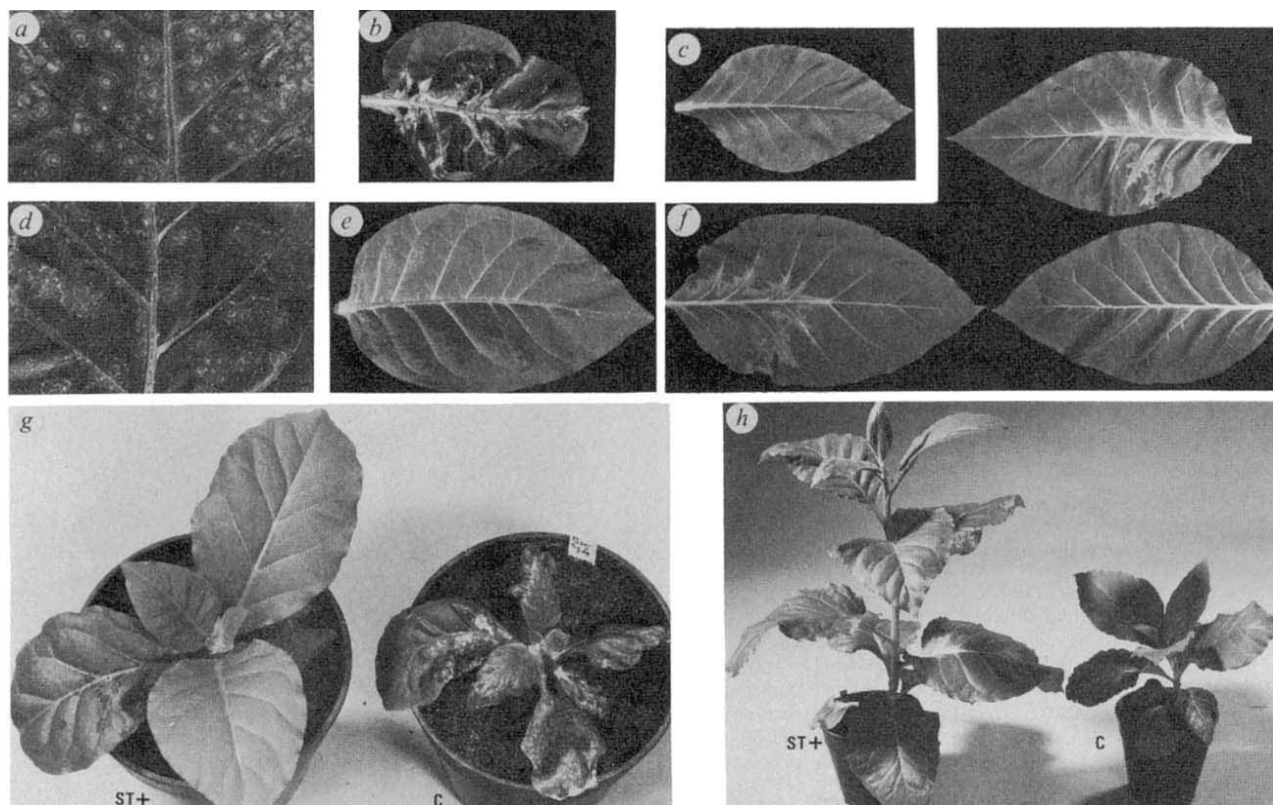


Fig. 3 Symptom development on *N. tabacum* var. Samsun non-transgenic control plants and transgenic plants containing an active ST+ gene construction at specific intervals after mechanical inoculation with TobRV. *a-c*, Leaf symptoms on non-transgenic control plants showing *a*, individual necrotic local lesions, 1 week after inoculation; *b*, strong systemic lesions and necrosis, 3 weeks after inoculation; *c*, mottled, small systemically infected leaves, 6 weeks after inoculation. *d-f*, Leaf symptoms on transgenic plants containing ST+: *d*, local lesions, 1 week after inoculation; *e*, symptomless leaf, 3 weeks after inoculation; *f*, symptoms, ranging from weak systemic reactions without necrosis to symptomless, from leaves 6 weeks after inoculation. *g-h*, Growth of non-transgenic plants (marked C) compared with that of ST+ transgenic plants ('ST+'), 3 weeks (*g*) and 6 weeks (*h*) after inoculation.

Methods. Seed progeny from regenerated transgenic plants and nontransgenic controls were germinated on damp vermiculite at 20 °C. After 3 weeks they were transferred to 13-cm pots containing loam and grown in a biosafety containment glass house (23 °C day, 15 °C night, natural daylength and lighting) for a further week. Inoculum was obtained from leaves of cucumber var. Spacemaster (Krempins Seeds, Armidale, Australia) mechanically inoculated with TobRV, and heavily infected as judged by severe stunting ~100 lesions per inoculated leaf. That the virus was free of satellite RNA was verified by nucleic acid hybridization of extracts from the cucumber plants. Fresh leaf samples (5 g) were finely ground in 10 ml of 0.1 M sodium phosphate buffer pH 7.0 at 4 °C, and used to infect the cotyledon leaves and expanded first true leaves of tobacco seedlings (100 µl per leaf) by mechanical inoculation using carborundum as abrasive. Following inoculation the leaves were rinsed with water and the seedlings grown under continuous light with 12 h 18–24 °C temperature cycles for four days. Seedlings were then returned to the biosafety containment glasshouse and monitored for symptom development.

symptom severity in transgenic plants which contained the ST+ and ST- gene constructions. Symptom severity was scored at 1, 3 and 6 weeks after inoculation of the seedlings with virus. Symptom development was similar for untransformed plants and those containing the SH+ gene construction. One week after inoculation, characteristic ringspot-shaped local lesions with severe necrotic centres were evident (Fig. 3*a*). Three weeks after inoculation, these plants showed severe systemic symptoms on all newly developed leaves (Fig. 3*b*) and stunting of the plants (Fig. 3*g*); after six weeks newly developed leaves were reduced in size and showed mottled symptoms characteristic of severe systemic TobRV infection (Fig. 3*c*). In all plants containing the ST+ gene construction, the primary lesions developed 1–2 days later than for nontransgenic plants and the centres of the lesions remained green, without necrosis (Fig. 3*d*); after three weeks new leaves seemed symptomless (Fig. 3*e*), with a mild systemic reaction only developing on some leaves after 5–6 weeks (Fig. 3*f*). At these later stages, leaf size was normal and plant growth was more vigorous than for the nontransgenic plants (Fig. 3*g* and *h*), and was similar to that of mock inoculated controls. The protection was maintained throughout the life of the ST+ plants, with flowering occurring when the plants were 10 weeks old.

In the ST+ transgenic plants the decreased symptoms are correlated with a lower level of replication of infectious virus compared with non-transgenic plants. When a hybridization probe prepared from cDNA of TobRV genomic RNA1 is used to assay the level of replicated virus genome in mechanically inoculated leaves the hybridization signal from ST+ plants is lower than that for non-transgenic controls (Fig. 4*a*). Furthermore, in the protected leaves higher on the ST+ plants there is no detectable virus, as judged by genome hybridizations, whereas it is present in the leaves of the systemically infected Samsun controls (Fig. 4*b,c*). The only detectable virus in the upper leaves of the ST+ plants occurs in the few leaves which show weak systemic symptoms six weeks after inoculation (Fig. 3*f*). Even there it is only present in the basal regions of those leaves and is associated with amplification of satellite RNA.

Disease development in the plants containing the ST- gene construction was different, with the initial ringspot lesions (one week after inoculation) possessing necrotic centres similar to infected nontransgenic controls. However, later lesions and subsequent symptoms at three and six weeks were more like those of the ST+ plants and, as these plants progressed, they developed a resistant phenotype and their growth rate recovered. The amplification of both (+) and (–)STobRV RNA in the ST+ and

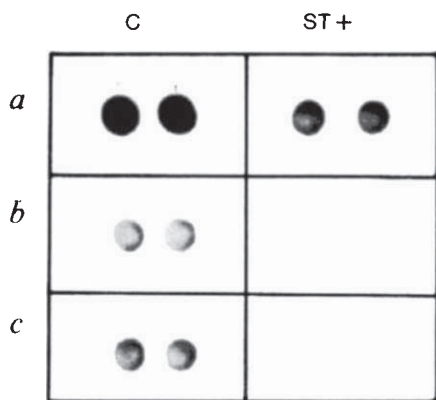


Fig. 4 Relative virus levels in leaves of an ST+ transgenic plant compared with a *N. tabacum* var. Samsun non-transgenic control plant four weeks after inoculation with TobRV. The autoradiographs show exposures of duplicate dots from equivalent leaf extracts following hybridization with radioactive probe complementary to TobRV virus genome. ST+, transgenic plant; C, non-transgenic plant. *a*, Leaf which was mechanically inoculated with virus, *b*, fourth leaf above inoculated leaf, *c*, eighth leaf above inoculated leaf.

Methods. Duplicate 20-mm-diameter disks were cut from leaves of ST+ transgenic and Samsun non-transgenic plants four weeks after infection by mechanical inoculation with TobRV as described in Fig. 3 legend. They were extracted by grinding in 1 ml 0.1 M sodium phosphate buffer pH 7.0, followed by shaking with 1 ml chloroform. After centrifugation to separate the phases, 3 μ l of the aqueous extract was spotted in duplicate on to nitrocellulose membrane and baked at 80 °C for two hours. The filter was hybridized according to Waterhouse *et al.*²⁷ except that 10^7 c.p.m. of nick-translated pTB9 plasmid was used as probe. Plasmid pTB9 contains a 1.7-kb cDNA insert derived from TobRV RNA1 (from Mark Anderson, DSIR, New Zealand). X-ray film (Kodak AR) was exposed to the filter for 48 h at -80 °C with intensifying screen.

ST- (Fig. 2*b*) correlated with the resistant phenotype. Although this might have been expected for the ST+ transformed plants, the result was surprising for the ST- plants because *in vitro* transcripts of the complementary strand previously were not biologically active when inoculated onto plants with virus¹³. However, we cannot exclude the possibility of transcription of the ST- gene from an adjacent plant promoter, or of transcription of (-)strand RNA by host RNA-dependent RNA polymerases¹⁴, to produce low levels of (+)strand STobRV sequences which were undetectable by RNA blot analysis of transgenic plants.

The production of novel virus resistance genes based on coat protein gene sequences of plant viruses has been shown for tobacco mosaic virus and alfalfa mosaic virus^{15,16}. This paper demonstrates that sequences of a satellite RNA of tobacco ringspot virus can be used to construct genes which, when incorporated into plants, confer a resistance phenotype on the plant when it is challenged with tobacco ringspot virus. That this approach to the synthesis of resistance genes may be more generally applicable to other plant viruses is shown by the recent production of genetic resistance to a different plant virus, cucumber mosaic virus¹⁷, using a satellite RNA unrelated to that of TobRV. A new approach to the synthesis of plant virus resistance genes is therefore now available.

We thank Drs John Randles and Richard Francki for the TobRV strain, Chris Howes for technical assistance; and Dr Paul Chu for advice on efficient mechanical inoculation of plants. J.H. was supported by CSIRO and Queen Elizabeth II postdoctoral fellowships. Experiments were carried out according to the guidelines of the Recombinant DNA Monitoring Committee, Department of Industry Technology and Commerce, Australia.

Received 17 June; accepted 12 August 1987.

1. Stace-Smith, R. *CMI/AAB Descr. Pl. Viruses No. 17* (1970).
2. Murrant, A. F. & Taylor, M. J. *gen. Virol.* **41**, 53-61 (1978).
3. Murrant, A. F., Taylor, M. J. & Duncan, G. H. *A. Rep. Scott. Hort. Res. Inst.* 1978, 98-99 (1979).
4. Schneider, I. R. *Science* **166**, 1627-1629 (1969).
5. Schneider, I. R. *Virology* **45**, 108-122 (1971).
6. Buzayan, J. M. *et al. Virology* **151**, 186-199 (1986).
7. Kiefer, M. C., Daubert, S. D., Schneider, I. R. & Bruening, G. *Virology* **121**, 262-273 (1982).
8. Hutchins, C. J. *et al. Pl. molec. Biol.* **4**, 293-304 (1985).
9. Forster, A. C. & Symons, R. H. *Cell* **49**, 211-220 (1987).
10. Prody, G. A. *et al. Science* **231**, 1577-1580 (1986).
11. Buzayan, J. M., Gerlach, W. L. & Bruening, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8859-8862 (1986).
12. Buzayan, J. M., Gerlach, W. L. & Bruening, G. *Nature* **323**, 349-353 (1986).
13. Gerlach, W. L., Buzayan, J. M., Schneider, I. R. & Bruening, G. *Virology* **150**, 172-185 (1986).
14. Fraenkel-Conrat, H. *Trends biochem. Sci.* **294**, 184-186 (1979).
15. Abel, P. P. *et al. Science* **232**, 738-743 (1986).
16. Tumer, N. E. *et al. EMBO J.* **6**, 1181-1188 (1987).
17. Hamison, B. D., Mayo, M. A. & Baulcombe, D. C. *Nature* **328**, 799-802 (1987).
18. Herrera-Estrella, L. *et al. EMBO J.* **2**, 987-995 (1983).
19. Walker, J., Howard, E. A., Dennis, E. S. & Peacock, W. J. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
20. An, G. *et al. EMBO J.* **4**, 277-284 (1985).
21. Hoekema, A., Hirsch, P. R., Hooykaas, P. J. & Schilperoort, R. A. *Nature* **303**, 209-213 (1983).
22. Ditta, G., Stanfield, S., Corbin, D. & Helinski, D. R. *Proc. natn. Acad. Sci. U.S.A.* **77**, 7347-7351 (1980).
23. Tran Thanh Van, K., Toubart, P. & Cousson, A. *Nature* **314**, 615-617 (1985).
24. Ellis, J. G., Llewellyn, D. J., Dennis, E. S. & Peacock, W. J. *EMBO J.* **6**, 11-16 (1987).
25. Palukaitis, P., Garcia-Arsenal, F., Sulzinski, M. A. & Zaitlin, M. *Virology* **131**, 533-545 (1983).
26. Boyer, P. D. *Nucleic Acids Res.* **14**, 7505 (1986).
27. Waterhouse, P. M., Gerlach, W. L. & Miller, W. A. *J. gen. Virol.* **67**, 1273-1281 (1986).

The role of clonal selection and somatic mutation in autoimmunity

Mark J. Shlomchik*, Ann Marshak-Rothstein†, Claudia B. Wolfowicz†, Thomas L. Rothstein†‡ & Martin G. Weigert*

* Institute for Cancer Research, Fox Chase Cancer Center, Philadelphia, Pennsylvania 19111, USA
 † Departments of † Microbiology and ‡ Medicine, Boston University Medical Center, Boston, Massachusetts 02215, USA

Polyclonal activation has been proposed as the reason that autoantibodies are produced during autoimmune disease¹⁻³. This model denies a role for specific antigen selection of B cells and predicts instead a multiclonal population of unmutated or randomly mutated autoantibodies. We have found that the genetic features and clonal composition of spontaneously derived immunoglobulin G (IgG) anti-self-IgG (rheumatoid factor (RF)) autoantibodies derived from the autoimmune MRL/*lpr* mouse strain are inconsistent with both the predictions of this model and the actual outcome of experimental polyclonal activation^{4,5}. Instead we have found that MRL/*lpr* RFs are oligoclonal or even monoclonal in origin. They harbour numerous somatic mutations which are distributed in a way that suggests immunoglobulin-receptor-dependent selection of these mutations. In this sense, the MRL/*lpr* RFs resemble antibodies elicited by exogenous antigens after secondary immunization⁶⁻⁸. The parallels suggest that, like secondary immune responses, antigen stimulation is important in the generation of MRL/*lpr* RFs.

The MRL/*lpr* and other mouse strains which carry the autosomal recessive mutation *lpr* (ref. 9), are models of autoimmune disease¹⁰⁻¹³. The types of autoantibodies they produce bear a striking resemblance to those found in human systemic lupus erythematosus and rheumatoid arthritis patients^{9,14-16}. One model to account for such autoantibodies suggests that self-specific antibodies are encoded by the germline V gene repertoire and are thus present in both normal and diseased individuals. In normal mice, self-reactive cells are suppressed, but in disease, an abnormal polyclonal B-cell activation occurs resulting in the pathological secretion of autoantibodies^{1-3,17}.

Table 1 Summary of RF hybridoma characteristics

Cell name	Mouse*	Clone name†	Isotype	V _κ group‡	J _κ	V _H family	D _H	J _H	CDR3 length¶
AM1	4		IgM	4	2 ⁸	J558	N/A	4	10
AM2	1	A	IgG ₁	10	4	J558	SP2.8'	4	10
AM3	1	A	IgG ₁	10	4	J558	Sp2.8'	4	10
AM4	1	A	IgG ₁	10	4	J558	SP2.8'	4	10
AM5	1	A	IgG ₁	10	4	J558	SP2.8'	4	10
AM6	4		IgG ₁	1	1	J558	SP2.2	4	14
AM7	4	D	IgG ₁	21	1	Sm7	SP2	4	14
AM8	4	D	IgG ₁	21	1	Sm7	SP2	4	14
AM9	4		IgG _{2b}	21	2'	J558	SP2	4	8
AM10	4		IgG _{2b}	21	1	J558	SP2	4	9
AM11	4		IgG _{2b}	1	1	J558	SP2	2	12
AM12	4		IgG _{2b}	21	2'	J558	N/A	4	12
AM13	4	E	IgA	8	4	J558	SP2	3	11
AM14	4	E	IgA	8	4	J558	SP2	3	11
AM15	4		IgA	21	1	J558	FL16	4	9
AM16	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM17	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM18	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM19	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM20	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM21	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM22	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM23	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM24	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM25	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM26	2	B	IgG ₃	19	4	Vgam3.8	SP2	4	13
AM28	3	C	IgA	24	1	J558	SP2	4	12
AM29	3	C	IgA	24	1	J558	SP2	4	12
AM30	3	C	IgA	24	1	J558	SP2	4	12
AM31	3	C	IgA	24	1	J558	SP2	4	12

For each cell line, the mouse from which it was derived, the heavy-chain isotype and a summary of the types of V_H, D_H, J_H, V_κ and J_κ genes which it expresses are given. The sera of mouse no. 2 and mouse no. 3 were tested for the heavy-chain isotypes of RF antibody; the dominant serum RF isotype matched the isotype found among the hybridomas in both cases (C.B.W. *et al.*, manuscript in preparation). Twenty-six of these 31 hybridomas were derived from MRL/*lpr* or MRL/*lpr*. IgH^c/4 mice. Five hybridomas were derived from C3H/*lpr*. Because the latter five are similar in sequence characteristics to the 26 others, we will refer to the collection as MRL/*lpr*-derived. The methods and results of assaying these fusions and the application of the assay to *lpr*-derived serum samples will be described in detail elsewhere (C.B.W. *et al.*, manuscript in preparation). Briefly, hybridomas were derived from four separate fusions, each using splenic B cells from an individual mouse for hybridization with the SP2/O myeloma partner. Fusions were performed by standard methods³⁶, except that B cells from a single spleen were divided into 4–5 aliquots and separate mini-fusions were performed. Cells were diluted and plated into 96-well plates such that plates which contained 1–2 hybrids per well could be screened for RF activity. Because hybridomas are derived from fusions of different aliquots of cells from the same spleen and/or were identified in different wells, re-isolation of the same hybridomas is precluded. At 10–14 days after plating, aliquots of supernatants from individual wells were assayed for RF activity using a modification of standard RF solid-phase plate assays designed to detect IgM-RF (ref. 37). Standard assays bind IgG-κ monoclonal antibodies directly to plastic plates as targets, block the nonspecific binding sites on the plates with bovine serum albumin (BSA), and incubate with test supernatants or serum samples. After extensive washing, IgM anti-IgG activity is detected by revealing bound IgM using a IgM-specific labelled antiserum. Briefly, our modification is to use IgG/λ monoclonal antibodies as targets. These are bound to plates indirectly through their specificity for the hapten 4-hydroxy-3-nitrophenyl coupled to BSA. After blocking, incubation and washing, bound RF is revealed by a labelled anti-κ reagent, allowing the detection of IgG/κ-RF as well as IgM/κ RF. That this assay detects IgM-RF was confirmed by testing on a representative panel of IgM-RF identified in other laboratories using the standard assay. Cells from wells identified by this screen were cloned at least once by limiting dilution and the isotype of the RF was determined by developing the same assay with labelled isotype-specific reagents (Southern Biotechnology).

* Mice nos 1 and 2 were MRL/*lpr*.IgH^c/4 (C.B.W. *et al.*, manuscript in preparation) mouse 3 was C3H/*lpr* (IgHⁱ) and mouse 4 was MRL/*lpr* (IgH^j). All mice were bred and maintained at the Boston University School of Medicine animal colony. Original MRL/*lpr* and C3H/*lpr* breeding pairs were obtained from Jackson Laboratories. The number of hybridomas screened (estimated from the number of wells screened and the average number of independent hybrids per well) and the percentage of RF-positive hybrids for mice nos 1 to 4 are: 1,000 hybrids, 17% RF positive; 100 hybrids, 13% RF positive; 300 hybrids, 21% RF positive; and 700 hybrids, 9% RF positive.

† Clone assignments are made as described in the text. Cell lines with no other clonally related line are not given clone names.

‡ Assignments of sequences to homology groups are made as described^{4,5}, based on >80% sequence homology to known prototypes. The nomenclature for V_κ group numbers is as described by Potter *et al.*³⁸, and for V_H families by Brodeur and Riblet³⁹, except that Vgam3.8 and Sm7 are the names given to families of V_H genes identified by Winter *et al.*⁴⁰ and Shlomchik *et al.*⁵, respectively.

§ The prime symbol (') indicates that this gene is probably the MRL allele of a known BALB/c gene. This conclusion is based on the similarity to the BALB/c gene and finding the same variation in several independent clones.

|| D_H designations are by homology to the closest known BALB/c gene⁴¹ as determined by eye. Where specific homologous D_H genes could not be identified, the D_H homology family which the gene most closely resembled is given. NA, not assignable: no significant resemblance could be found.

¶ CDR3 length is defined as the number of amino acids between the last amino acid encoded by V_H genes (two residues after the invariant cysteine) and the invariant tryptophan residue encoded by all J_H genes.

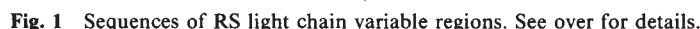


Fig. 1 (See previous page.) Nucleotide sequences of variable regions of RF light chains. Sequences were determined as described⁹, except that an additional oligonucleotide (5' AGCACC GAACGTGAGGGG 3') was used to determine the sequences shown in *b*. *a*, *b*, *c*, *d*, Hybridomas from mice 1, 2, 3 and 4 respectively. Groups of homologous sequences are compared to prototypes and identities indicated with dashes. In *d*, the different groups are each separated by a blank line and all groups are aligned by introducing blank spaces so that homologous residues correspond to the first sequence (AM13). The translations of the prototype sequences are given above the nucleotide sequences. In cases where more than two hybridomas have identical sequences, their names are listed to the left of the sequence. Nucleotides are numbered beginning with the first amino acid of the predicted mature peptide. In all cases, sequence was determined 5' of this position (not shown); usually, the sequence of all of the predicted leader peptide was determined. In some cases, additional sequence was also determined beyond the furthest 3' base shown. In *b*, a small portion of the 3'-most sequence of AM16, AM23, AM25 and AM26 was not determined; the points at which these sequences stop are indicated with the symbols *, +, ~ and ○ respectively. The functional domains defined by Kabat *et al.*²⁵ are indicated with either FR (framework region) or HVR (hypervariable region) and the corresponding number.

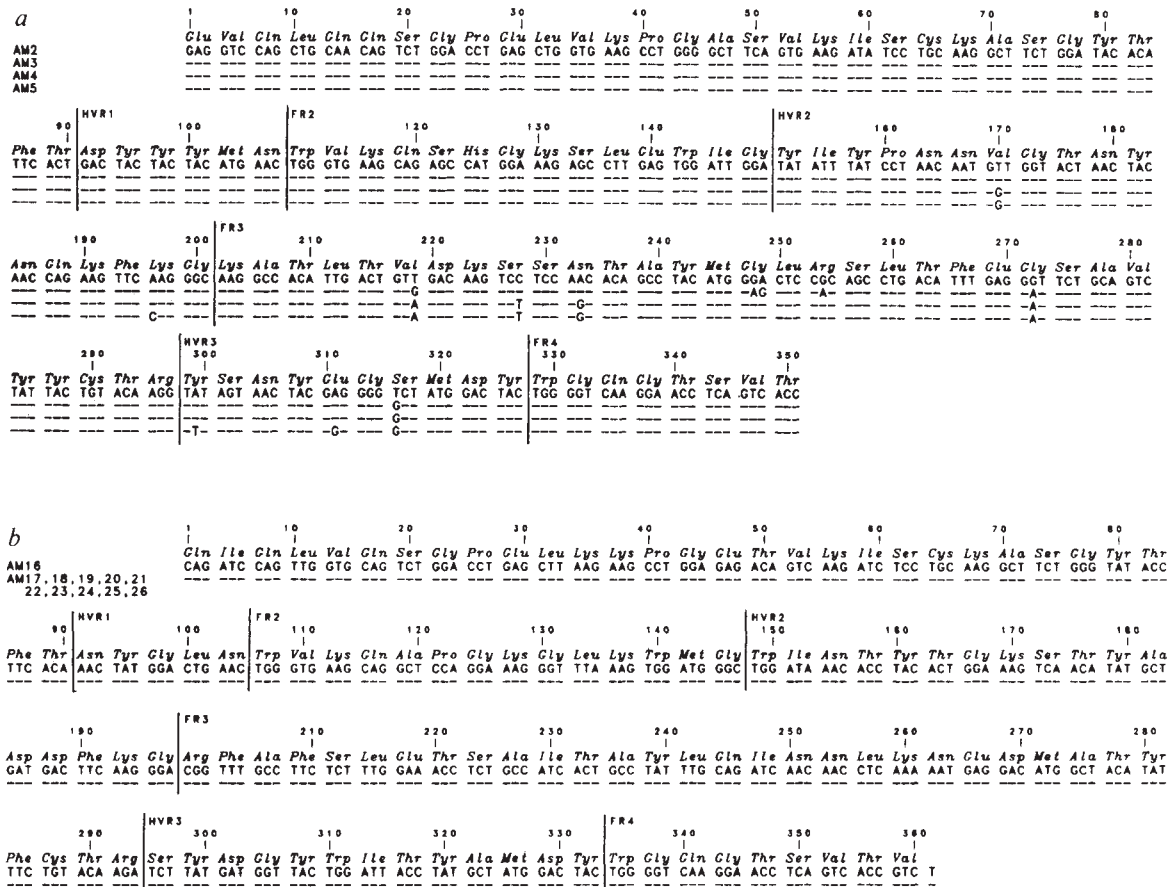


Fig. 2 (Above and opposite.) Nucleotide sequences of variable regions of RF heavy chains. Sequences are displayed in the same order and format as Fig. 1. In *d*, only a partial sequence for AM15 is shown; sequence not determined is indicated by spaces enclosed by brackets.

This model predicts that autoantibodies would resemble those derived after polyclonal activation of normal B cells. A second model invokes self-antigens, not polyclonal activation, as the driving force for the production of autoantibodies¹⁸⁻²¹. This 'antigen driven' model, predicts that autoantibodies would resemble antibodies produced during the secondary immune response to foreign antigens; these are usually derived from clonally related precursors, somatically mutated (often reflecting selection of mutations by antigen), and isotype switched^{6,7}.

To distinguish between the models, we have constructed a panel of 31 hybridomas which secrete autoantibodies against immunoglobulin IgG_{2a} (a specificity known as rheumatoid factor (RF)). As the models make different predictions about the nature of the variable (V)-region sequences of such autoantibodies, we have determined their sequences and compared these sequences to sequences of RFs derived from polyclonal activation^{4,5}.

The sequences reveal a striking feature of these hybridomas: there are multiple independent isolates which derive from the same B-cell precursor. Our criteria for establishing such clonal

relatedness are identity at the somatically formed V_H-D_H , D_H-J_H , and V_K-J_K junctions and use of the same sets of V_H , D_H , J_H , V_K and J_K genes (see Figs 1 and 2 for sequences and Table 1 for clone designations). The context of rearrangements of non-productive alleles at the immunoglobulin loci, as detected by Southern blotting, supports the clone assignments (data not shown).

Three of the four mice (nos 1, 2 and 3) yielded hybridomas which derived from just one B-cell precursor each (clones A, B, and C). The fourth mouse was more clonally diverse, although it too had clones of two or more members (clones D and E). This result is different from that obtained from eight previous fusions of normal mouse spleen cells (either after polyclonal activation or primary immune complex stimulation) from which more than one RF-producing hybridoma was isolated; this survey did not reveal any clonally related hybridomas among 42 tested^{4,5}. Thus, in these *lpr* mice, RF-secreting cells are derived from significantly fewer precursors than the RFs derived from normal mice.

[illegible]

These V-region sequences (Figs 1 and 2) also contain many somatic mutations. For example, among the members of clone A there are numerous single-base-pair differences. Such nucleotide differences must have been acquired somatically during clonal outgrowth. Comparison of hybridoma V-region sequences to highly similar germline or consensus sequences has also been used to identify mutations. For example, the V_{κ} sequences of AM6 and AM11 are highly homologous to each other and the germline gene designated $V_{\kappa}1C$ (ref. 22). The restriction maps of the light-chain rearrangements in these lines are also consistent with their using this V_{κ} gene (data not shown). The V_{κ} sequences of AM7, 8, 9, 10 and 15 and the V_H sequences of AM9, 10 and 12 have been surveyed for mutation in a similar way.

The number of mutations per V region analysed was 4.4, with the most mutated V region having seven mutations (clone B was excluded from this calculation, see below). Because differences from the germline sequences which are shared by all sequenced members of a clone are not detected by this analysis, this

is a minimum estimate. Nonetheless, the content of mutations is significantly higher than that found in our previous studies of RF from normal mice. The latter have only 0.5 mutations per V region, with the most mutated V region having two mutations (data in refs 4 and 5).

These RF *V* regions have an unexpectedly high ratio of replacement (R) to silent (S) mutations ('R:S ratio') and a non-random distribution of mutations in the *V* regions, which argues that antigen has positively selected B cells during the outgrowth of these RF clones. Random mutations in regions of a protein which need not be conserved to maintain function should yield an R:S ratio of 2.9 (ref. 23). In regions such as framework regions (FRs) of antibodies which need to be conserved to maintain function, this ratio is expected to be lower and is about 1.5, based on the extent of conservation seen in comparisons of diverse *V*-region sequences^{6,24,25}. On the other hand, the ratio in hypervariable regions (HVRs) of antibodies might approach 2.9, as there are relatively fewer conserved residues in HVRs. But ratios significantly higher than 2.9 can only be

achieved through positive selection of R mutations. For example, if a R mutation which increases the affinity of an antibody for antigen confers a selective advantage to a mutated B cell (that is a higher probability to divide compared to siblings), then this R mutation will be fixed and propagated in the antigen-specific B cell population. Most of such mutations are expected to be in the HVRs, since these are thought to include most of the antigen contact residues^{26,27}. The mutations found in each of the clones analysed show evidence of selection for R mutations in HVRs and against R mutations in FRs. In clone A, the R:S ratio is 9.0 for the HVRs and 1.2 for the FRs, whereas the light chains of AM9, AM10 and AM11 have two, one and three R mutations in HVR and no S mutations. The R:S ratio in HVRs of all sequences analysed is 7.7 and for FRs is 1.4. This pattern of somatic mutations in the V regions of these clones can only be explained by positive selection of certain B cells on the basis of their antigen receptors, presumably because these mutated receptors provide superior ability to bind antigen.

The combination of shared and unique mutations in clone A leads to a genealogy (Fig. 3a). In this regard, clone A resembles previously reported clones derived from secondary immune responses^{6,7}. However, clones B, C, D and E (see Fig. 3 for genealogies) are different from clone A because they contain multiple members which are identical in sequence. In clone C, four of five members have identical sequences. But the most striking example of this pattern is clone B, which has 11 identical members, implying that the clone of B cells from which these hybridomas were derived had undergone, just before isolation, a large number of divisions without any mutation occurring.

Clones B and C could not have been mutating at the rate found for anti-influenza haemagglutinin clones (10^{-3} per base pair (bp) per division)²⁸, as this rate predicts that after each division, every other sibling will possess a new mutation. Indeed, the maximum mutation rate for clone B is much lower: in the unlikely event that the 11 cells we studied represent all of the progeny of the clone and that it had gone through the minimum number of generations needed to create these progeny, the rate would have to be $<10^{-4}$ per bp per division to have $<5\%$ chance of creating 11 identical cells. However, the efficiency of rescue of activated cells by the hybridoma method is low, around 1/1,000 to 1/10,000 (refs 29 and 30). It is more reasonable to assume that these 11 identical hybrids are just a sample of the activated cells in the mouse at the time of immortalization and that the number of identical members of the clone must have been quite large. Conservatively, if 1/1,000 cells from the spleen of this mouse were captured by the fusion process, then the mutation rate in this clone must have been $<5 \times 10^{-8}$ for the 13 divisions needed to create the minimum estimated 11,000 identical B cells in the clone.

Mutation at a high rate may never have started in this clone, or it may have started and then ceased at some point during clonal outgrowth. Although proof of either alternative awaits cloning of the relevant germline genes, the idea that mutation started then stopped is supported both by the precedent from clones of other specificities^{8,31}, which contain identical examples which had undergone somatic mutation, and clone C, which has four identical members and one different member, showing that mutation had occurred in at least one growing branch. Furthermore, comparison with highly homologous sequences from the V_k group expressed in clone C, represented in Fig. 1c by the prototype A15 (ref. 5), strongly suggests that the precursor of the four identical cell lines itself had numerous mutations.

Because, in these fusion experiments, 10–20% of all hybridomas were making RF, and RFs in these mice generally derived from just one precursor B cell per mouse, the RF precursors must have undergone great clonal expansion. This domination of single clones must be the result of the selective advantage of certain (rare) B cells. What is the reason for this

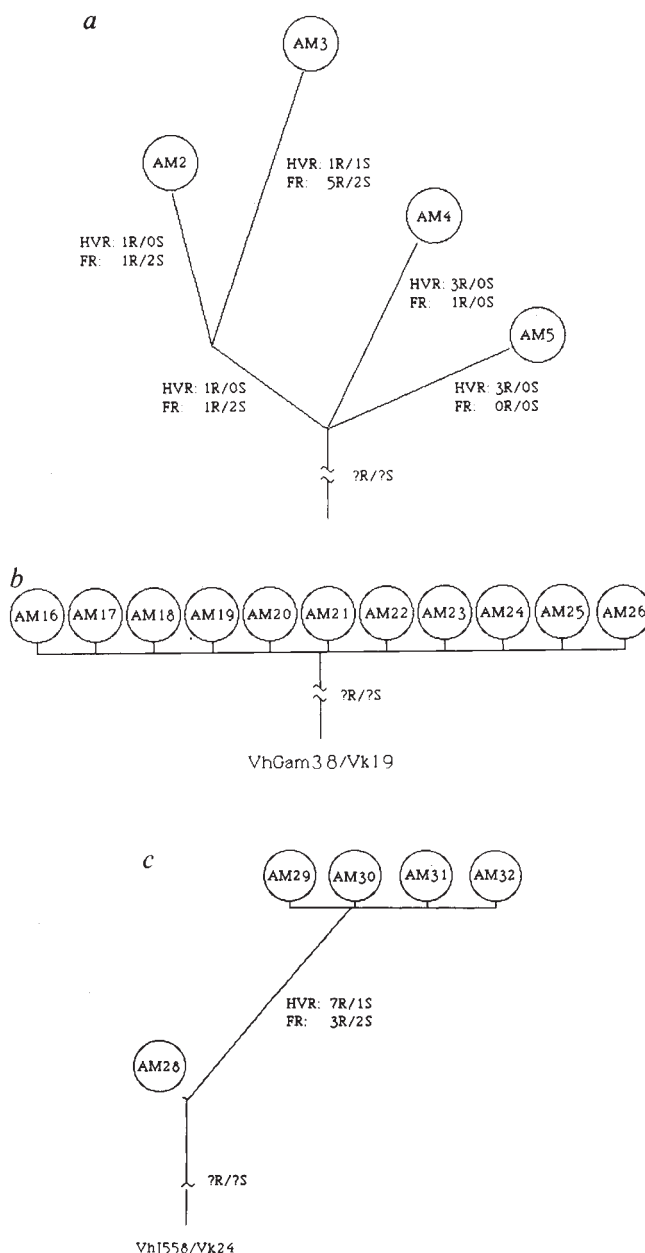


Fig. 3 Genealogies of clones A, B and C, represented in a, b and c respectively. Genealogies are constructed from the pattern of shared and unique mutations in a clone, assuming that shared mutations represent single events and not independent parallel mutations. The lengths of branches are roughly proportional to the number of mutations on the branches, except that for clarity, short lines are used to separate identical cell lines in b and c. Near each branch is an indication of the location (either HVR, hypervariable region or FR, framework region), number and type (either R, replacement or S, silent) of mutations found in that branch.

selective advantage? It is not simply the expression of a certain set of V genes, as evidenced by the differences between clones. One reason, as suggested by the high R:S ratios, is mutations which confer better binding to antigen. A further advantage is provided to clones derived from cells which are no longer mutating at a high rate because they will undergo greater expansion than clones in which cells are still mutating at high rates. This is because a large proportion of mutations would be expected to eliminate affected cells from the antigen-specific pool either by destroying the structure of the immunoglobulin

molecule or by eliminating the binding affinity of the antibody for the original antigen. Thus, mono- or oligo-clonal dominance and receptor-mediated selection are consistent with the idea that MRL/*lpr* RFs are the result of chronic stimulation of B cells with an antigen (most probably IgG-containing immune complexes^{32,33}), but are not predicted by polyclonal activation models.

The clonal evolution of RFs can be related to the natural history of autoimmunity in the MRL/*lpr* mouse. Disease manifestations and some types of autoantibodies (including RF and anti-Sm) are not present in the serum of young MRL/*lpr* mice, but appear dramatically and suddenly in individual mice as they age^{34,35}. The rapid rise in titres could correspond to the 'success' of one or a few clones which had acquired mutations that improved affinity, encountered antigen, grew with selective advantage, and perhaps finally lost the capacity to mutate, so that they achieved even faster clonal growth rates.

We thank Anita Cywinski and Sharon Lyons for technical work, Donna Hnosko for preparation of the manuscript, Sam Litwin for statistical analysis, P. Felix and J. Felix for help with cells, and S. Clarke, R. Eisenberg, D. Nemazee, A. Haberman and D. Pisetsky for critical reading of the manuscript. This work was supported by grants from the NIH to M.W., T.R. and A.M. and an appropriation from the Commonwealth of Pennsylvania. M.S. is a trainee of the Medical Scientist Training Program at the University of Pennsylvania.

Received 23 March; accepted 10 June 1987.

1. Hang, L. *et al. J. exp. Med.* **157**, 874-883 (1983).
2. Dziarski, R. *J. Immun.* **128**, 1018-1025 (1982).

3. Teodorescu, M. *Immun. Rev.* **55**, 155-178 (1981).
4. Shlomchik, M. J. *et al. J. exp. Med.* **164**, 407-427 (1986).
5. Shlomchik, M. J., Nemazee, D. A., Van Snick, J. & Weigert, M. G. *J. exp. Med.* **165**, 970-987 (1987).
6. Clarke, S. H. *et al. J. exp. Med.* **161**, 687-704 (1985).
7. Caton, A. J., Brownlee, G. G., Staudt, L. M. & Gerhard, W. *EMBO J.* **5**, 1577-1587 (1986).
8. Siekevitz, M., Kocks, C., Rajewsky, K. & Dildrop, R. *Cell* **48**, 757-770 (1987).
9. Izui, S. *et al. J. Immun.* **133**, 227-233 (1984).
10. Theofilopoulos, A. N. & Dixon, F. J. *Adv. Immun.* **37**, 269-390 (1985).
11. Wofsy, D., Hardy, R. R. & Seaman, W. E. *J. Immun.* **132**, 2686-2689 (1984).
12. Hang, L., Theofilopoulos, A. N. & Dixon, F. J. *J. exp. Med.* **155**, 1690-1701 (1982).
13. Kelley, V. & Roths, J. B. *Clin. Immunopath.* **37**, 220-229 (1985).
14. Notman, D. D., Kurata, N. & Tan, E. N. *Ann. intern. Med.* **83**, 464-469 (1975).
15. Munthe, E. & Natvig, J. B. *Clin. exp. Immun.* **12**, 55-70 (1972).
16. Youinou, P. Y., Morrow, J. W., Lettin, A. W. F., Lydyard, P. M. & Roitt, I. M. *Scand. J. Immun.* **20**, 307-315 (1984).
17. Theofilopoulos, A. N., Shawler, D. L., Eisenberg, R. A. & Dixon, F. J. *J. exp. Med.* **151**, 446-466 (1980).
18. Portanova, J. P., Claman, H. N. & Kotzin, B. L. *J. Immun.* **135**, 3850-3856 (1985).
19. Shores, E. A., Eisenberg, R. A. & Cohen, P. L. *J. Immun.* **136**, 3662-3667 (1986).
20. Rappard-Van Der Veen, F. M. *J. Immun.* **132**, 1814-1819 (1984).
21. Nelson, J. L., Nardella, F. A., Opplinger, I. R. & Mannik, M. *J. Immun.* **138**, 1391-1396 (1987).
22. Corbet, S., Milili, M., Fougereau, M. & Schiff, C. *J. Immun.* **138**, 932-939 (1987).
23. Jukes, T. H. & Kling, J. L. *Nature* **281**, 605-606 (1979).
24. Wu, T. T. & Kabat, E. A. *J. exp. Med.* **132**, 211-250 (1970).
25. Kabat, E. A., Wu, T. T., Bilofsky, H., Reid-Miller, M. & Perry, H. *Sequences of Proteins of Immunological Interest* (US Govt Printing Office, Bethesda, Maryland, 1983).
26. Segal, D. *et al. Proc. Natn. Acad. Sci. U.S.A.* **71**, 4298-4302 (1974).
27. Amit, A. G., Mariuzza, R. A., Phillips, S. E. V. & Poljak, R. J. *Nature* **313**, 156-158 (1985).
28. McKean, D. K. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 3180-3186 (1984).
29. Claflin, J. L. & Williams, K. *Curr. Top. Micro. Immun.* **81**, 107-109 (1978).
30. Andersson, J. & Melchers, F. *Curr. Top. Micro. Immun.* **81**, 130-132 (1978).
31. Claflin, J. L., Berry, J., Flaherty, D. & Dunnick, W. *J. Immun.* **138**, 3060-3069 (1987).
32. Nemazee, D. *J. exp. Med.* **161**, 242-256 (1985).
33. Coulie, P. & Van Snick, J. *J. exp. Med.* **161**, 88-92 (1985).
34. Tarkowski, A., Czerlinsky, C. & Nilsson, L.-A. *Clin. exp. Immun.* **58**, 7-12 (1984).
35. Cohen, P. L. *et al. Cell. Immun.* **96**, 448-454 (1985).
36. Marshak-Rothstein, A. *et al. J. Immun.* **133**, 2491-2497 (1978).
37. Van Snick, J. L., Stassin, V. & De Lestre, B. *J. exp. Med.* **157**, 1006-1019.
38. Potter, M., Newell, J. B., Rudikoff, S. & Haber, E. *Molec. Immun.* **19**, 1619-1630 (1982).
39. Brodeur, P. H. & Riblet, R. *Eur. J. Immun.* **14**, 922-930 (1984).
40. Winter, E., Radbruch, A. & Krawinkel, U. *EMBO J.* **4**, 2861-2867 (1985).
41. Kurosawa, Y. & Tonegawa, S. *J. exp. Med.* **155**, 201-218 (1982).

Chemical structure of the morphogen differentiation inducing factor from *Dictyostelium discoideum*

Howard R. Morris*, Graham W. Taylor‡, Mark S. Masento*, Keith A. Jermyn† & Robert R. Kay‡

* Department of Biochemistry, Imperial College, London SW7 2AZ, UK

† Imperial Cancer Research Fund, Clare Hall Laboratories, South Mimms, Hertfordshire EN6 3LD, UK

‡ MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

Morphogens are signal molecules presumed to exist in embryos and to be involved in establishing the spatial pattern of cells during development. Differentiation inducing factor (DIF) has the properties of a morphogen required for producing the prestalk/prespore pattern in the aggregate formed by cells of the slime mould *Dictyostelium* in response to starvation. DIF-1, the major bioactive species after purification, has now been identified using a combined microchemical, spectroscopic and synthetic approach. The structure is defined as 1-(3,5-dichloro-2,6-dihydroxy-4-methoxyphenyl)-1-hexanone, and represents a new class of effector molecule. The availability of relatively large quantities of synthetic and isotopically labelled materials should now allow progress towards a detailed understanding of the pattern-forming processes in *Dictyostelium* development.

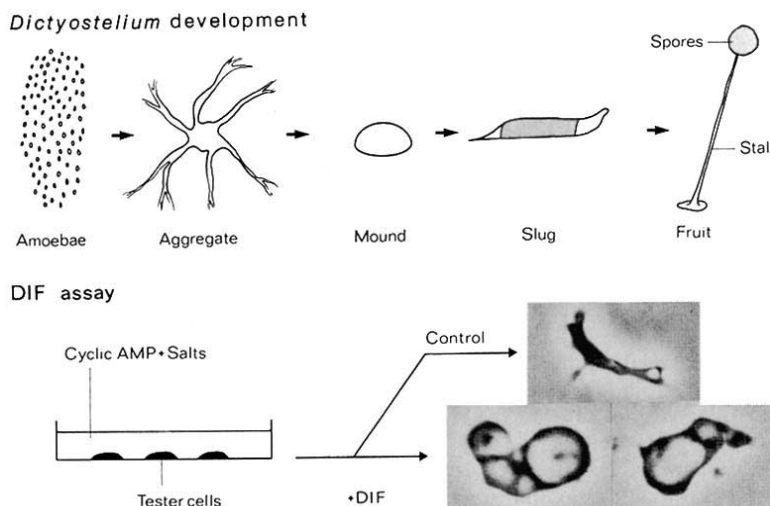
Experiments of more than 50 years ago on amphibian embryos showed that the fate of a developing cell could be totally altered

merely by transplanting it to a different part of the embryo. It is widely believed that this position-dependent development must reflect the existence of localized inductive signals in the embryo and it has remained a major challenge in developmental biology to understand this signalling process in any organism^{1,2}. One prerequisite is to identify the signals, or morphogens, themselves and recently two distinct strategies have shown promise. One strategy is to identify^{3,4} and clone genes affecting embryonic pattern. This could lead directly to the discovery of a protein morphogen, though some of the candidates to date, such as the *notch* and *caudal* genes of *Drosophila*^{5,6} and the *lin12* gene of *Caenorhabditis*⁷, could equally encode receptor or other proteins which mediate the activity of the true morphogens. In addition, this cloning strategy would be very difficult to apply to a non-protein morphogen as this could only be identified by reconstructing its biosynthetic pathway from the genes. An alternative strategy that avoids these problems is to use bioassays to detect morphogenetically active substances in the embryo. In this way, several morphogenetically active substances have been discovered, including *Hydra* head factor⁸, retinoic acid^{9,10}, the amphibian mesoderm-inducing factors¹¹ and DIF-1, the major differentiation inducing factor isolated from the slime mould *Dictyostelium discoideum*.

Starving *Dictyostelium* cells aggregate in response to relayed cyclic AMP signals to form small mounds, each of up to 10⁵ cells (Fig. 1, top panel). Such a mound lacks any obvious pattern but as it transforms into a cylindrical slug a simple pattern emerges, with prestalk cells differentiating in the anterior and prespores in the posterior. Later these precursor cell types differentiate into the stalk and spore cells of the mature fruiting body. The differentiation of amoebae into stalk and spore cells can be more conveniently studied using defined *in vitro* culture conditions, where morphogenesis is prevented. In this way DIF was discovered¹² as a dialysable factor which induces stalk cell differentiation amongst isolated amoebae¹³. Using the stalk-cell-induction bioassay for quantitation¹⁴ (Fig. 1, bottom panel), the

§ Present address: Department of Clinical Pharmacology, Royal Postgraduate Medical School, Du Cane Road, London W12 0HS, UK.

Fig. 1 Top, *Dictyostelium* development takes about 24 h to complete, is initiated by starvation and proceeds with very little cell division. A few hours after plating on a solid surface the amoebae aggregate toward collection points by chemotaxis to propagated waves of cAMP. The resultant cell masses split into mounds of about 10^5 cells which elongate upwards and then fall on their side to form migrating slugs. At this stage the prestalk/prespore pattern becomes apparent and various markers have been found to distinguish the two cell types (prespore and spore cells are shaded). After a variable period the slug ceases migration, rounds up and culminates to produce a fruiting body in which a cellular stalk supports a mass of refractile spores. Below, in the DIF assay, starving cells of a particularly responsive strain of *D. discoideum* called V12M2 are plated at low density (10^3 cm^{-2}) in tissue culture dishes under simple salts solution containing cAMP and antibiotics. In the absence of added DIF, the cells remain amoeboid and gradually get smaller. In the presence of DIF they start to vacuolate after 24 h and form mature stalk cells by 36 h. These are scored by phase contrast microscopy and in the range 10–50% stalk cells there is an approximately linear increase with DIF concentration^{14,15}. About 10^{-13} mol of DIF-1 can be detected in a 2-ml assay.



major activity released during development, DIF-1, has been purified by a combination of solvent extraction and reverse-phase HPLC (ref. 15). In this paper we describe the structure elucidation of DIF-1.

Although the amounts of endogenous DIF-1 initially available could not be quantified until the structure had been solved and synthesized, retrospective calculation based on the extinction coefficient and bioactivity show that the early preparations were in the low nanogram range, with the final largest preparation of DIF-1 corresponding to only 50 µg of material. These amounts explain our inability to obtain spectroscopic data on DIF-1 in either mass spectrometric or NMR experiments on early preparations.

Initial experiments on DIF-1 were therefore designed to attempt functional group analysis on picogram to nanogram quantities of material using strategies similar to those reported in the structure elucidation of leukotriene D_4 (SRS-A)¹⁶ and β -cell tropin¹⁷, in which destruction of biological activity is interpreted as chemical modification of suspected functional groups. This work confirmed and extended pilot studies¹⁵, and the most informative experiments showed: first, the absence of amino groups in DIF-1, in that there was no change in bioactivity on treatment with acetic anhydride in methanol; and second, the probable presence of one or more carbonyl functions, demonstrated by complete loss of activity upon borohydride reduction. These interpretations took account of other chemical data on the molecule, such as acid or alkali stability to discount other possible chemical classes such as lactone, amide or ester.

Further conclusions, albeit less certain, could be drawn from other microchemical derivatization experiments. Esterification with methanol/HCl resulted in no change in the bioactivity of DIF-1, implying the absence of carboxyl groups. However, treatment with diazomethane did abolish the bioactivity suggesting the presence of a phenolic or other acidic hydroxyl group, although the ready incorporation of methylenes from the reagent by, for example, the known ring expansion of cyclic ketones could not be discounted. Treatment of DIF-1 with acetic anhydride in pyridine (to acetylate hydroxyl functions) only reduced the DIF-1 bioactivity by some 50%, thus not resolving the problem of whether hydroxyl groups were present. DIF-1 was also stable to catalytic hydrogenation over platinum oxide at 2 atm, although diborane reduction resulted in inactivation.

Ultraviolet spectra of the subsequent scaled-up DIF-1 preparations were variable, showing broad absorbance with a maximum ($\lambda_{\max}$) between wavelengths 273 nm and 280 nm in different preparations. These data indicated the presence of a

relatively highly conjugated chromophore such as aromatic, triene or dienone, the latter idea accommodating the conclusion of carbonyl from the borohydride data. A proton NMR spectrum taken at this stage failed to produce any signal other than impurities, and that in itself was indicative of the need for scaled-up DIF preparations.

Analysis by fast atom bombardment mass spectrometry (MS) of a DIF-1 preparation gave no result. However, an electron impact experiment using a temperature-programmed source to distil off the accompanying impurities fractionally, produced the mass spectrum shown in Fig. 2, at a temperature of 130 °C. Significantly, this spectrum was observed at the same temperature in subsequent DIF-1 preparations and always corresponded to the elution position of DIF even when this was altered in the HPLC profile. Therefore, although the ultraviolet data did not allow a $\lambda_{\max}$ to be assigned to DIF at this stage, the mass spectrum provided the crucial means of approach to the structure of DIF, because it could be monitored, manipulated in derivatization experiments and ultimately interpreted to deduce the chemical identity of DIF-1.

The mass spectrum (Fig. 2, Table 1) could be partially interpreted as follows. A molecular ion (M^+) was apparent at m/z 306, defining for the first time the relative molecular mass of DIF-1. Isotopes associated with this ion at m/z 308 and 310 produced a cluster with a measured ratio of roughly 9:6:1, giving definitive evidence for the presence of two chlorine atoms in the DIF-1 structure. In other less active natural analogues of DIF we have discovered mono- and des-chloro structures.

Table 1 Elemental composition of the major ions in the electron impact mass spectrum of DIF-1

Ion (m/z)	Elemental composition				Measured mass	Fragment loss
	C	H	O	Cl		
306	13	16	4	2	306.0424	M^+
288	13	14	3	2	288.0320	$-H_2O$
263	10	9	4	2	262.9876	$-C_3H_7$
250	9	8	4	2	249.9798	$-C_4H_8$
235	8	5	4	2	234.9563	$-C_5H_{11}$
220	7	2	4	2	219.9329	$-C_6H_{14}$

Each of the $^{35}\text{Cl}_2$ -containing ions was manually mass measured against a perfluorokerosene reference. The assignments were confirmed for some ions by measurement of the $^{35}\text{Cl}^{37}\text{Cl}$ isotope. The measured masses of each of the ions (with the exception of m/z 220) were within 1 p.p.m. of the calculated mass.

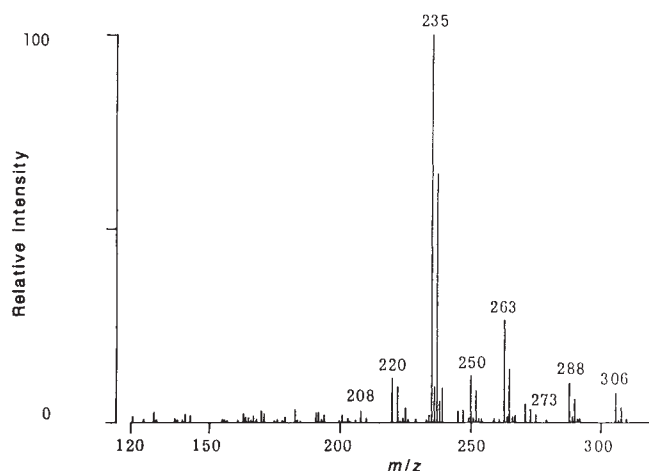


Fig. 2 The partial electron impact mass spectrum of DIF-1 at 70 eV, determined on an MS 50 mass spectrometer. The sample was fractionally distilled from a solid probe to remove impurities, over a temperature range of 50–300 °C. The spectrum assigned to DIF-1 appeared at 130 °C. The presence of two chlorine atoms in DIF-1 is defined by the 9:6:1 isotope ratio of the molecular ion (M^+ ; m/z 306, 308, 310) and each of the major fragment ions. High abundance fragment ions include m/z 288 ($M^+ - H_2O$), m/z 263 ($M^+ - C_3H_7$), m/z 235 ($M^+ - C_5H_{11}$) and m/z 250 ($M^+ - C_4H_8$, McLafferty rearrangement). Other ions correspond to further CH_3^+ losses (see, the MS-MS experiment). Note the absence of fragment ions below m/z 208, indicating the presence of a dichlorinated stable ring structure.

Accurate mass measurement at high resolution gave the exact mass of DIF-1 as 306.0424, corresponding to an atomic composition of $C_{13}H_{16}O_4Cl_2$, and showing the presence of five double-bond equivalents (rings, double bonds or both) in the DIF-1 structure. In a relatively small molecule this indicated a high degree of unsaturation and probable aromaticity. Ambiguities in the interpretation of the mass spectrometric fragmentation of DIF-1 were also resolved by accurate mass measurement at 10,000 resolving power (Table 1).

Further interpretation of the mass spectrum (to be reported in detail elsewhere), together with the other chemical data, led to the conclusion that DIF-1 possessed a normal- or iso-pentyl sidechain adjacent to a carbonyl group. Importantly, the lack of fragmentation below m/z 208 suggested the presence of a very stable nucleus in the DIF-1 structure of ~200 mass units, which contained the chlorine atoms because these were not lost as fragments. The absence of units such as $>CCl_2$ was defined by stability of DIF-1 (m/z 306) to base hydrolysis under conditions which destroyed chloramphenicol.

Microchemical derivatizations were repeated using the mass spectrum rather than the bioassay as a monitor. For example, reaction with diazomethane led to a shift in the m/z 306 molecular ion signal to m/z 334, corresponding to the addition of two methyl groups and indicating the presence of two acidic hydroxyl functions, most probably phenolic, in DIF-1. Attempts to block these groups with potassium carbonate/methyl iodide were not successful; however, difficulty in methylation was also encountered with experiments on standard polyhydroxy phenols under the same conditions.

The final (major) DIF-1 preparation of about 50 μ g was first used to produce a proton NMR spectrum, in a 24-h acquisition on a Bruker 250 MHz instrument. Residual protons in deuteromethanol provided reference signals. The spectrum was of poor quality, with intense absorbances between δ 1.2–1.7 p.p.m. masking many DIF-1 signals. A singlet at δ 3.96 p.p.m. and a perturbed triplet at δ 3.10 p.p.m. were interpreted to be an isolated methyl in an electronegative environment (aryl OMe), and a methylene alpha to a carbonyl, respectively. Spin decoupling of the δ 3.10 p.p.m. signal led to the collapse of an

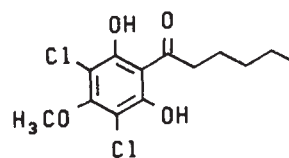


Fig. 3 The structure of the morphogen DIF-1 was determined by microchemical analysis, spectroscopy and synthesis to be 1-(3,5-dichloro-2,6-dihydroxy-4-methoxyphenyl)-1 hexanone.

estimated quintuplet at δ 1.67 p.p.m. to a triplet, confirming two of the supposed methylenes (one adjacent to carbonyl) deduced from the MS data. No protons were observed in the aromatic region of the spectrum, indicating that an aromatic ring would have to be fully substituted. In one of a series of spectra, a small signal was apparent at δ 10.32 (phenolic OH or other acidic proton) which disappeared on shaking in D_2O ; poor signal quality prevented a good integration of the relative number of protons in the spectrum. The production of a carbon-13 NMR spectrum which could have been invaluable to the study was not possible due to the small amount of DIF-1 available (at least a factor of ten less than required to produce a spectrum). Collisional activation MS-MS on a triple quadrupole (courtesy of W. Richter) using the dimethyl derivative of DIF-1, confirmed the proposed structural nucleus as a phenol by successive multiple losses of CO and CH_3^+ from a major fragment ion at m/z 263 ($M^+ - C_5H_{11}$).

Together, these data defined DIF-1 as a hexanophenone, with dichloro-dihydroxy-methoxy ring substitution. This can be manifested as 32 possible isomers (restricting the alkyl chain to either *n*- or *iso* from the MS data), and further structural information was required. Thus a programme of synthesis was started to determine the isomeric structure of DIF-1.

Ring oxygenation at 2,4,6 was chosen initially on putative biosynthetic grounds, and the various isomers synthesized. Starting from 1-methoxy-3,5-dihydroxy benzene, the hexanone function was attached by either the Fries or Hoesch reactions. Two isomers were generated, each of which was then chlorinated, purified by HPLC and fully characterized by ultraviolet, NMR and mass spectrometric examination. A number of other analogues of DIF-1 were also synthesized by this route, including the isopentyl isomer, C_3 – C_8 alkyl ketone homologues, and monochloro species. Comparison of the natural material with the synthetic dichloro- C_5 alkyl ketone isomers by HPLC, GC, MS, NMR and biological activity defined the structure of DIF-1, shown in Fig. 3, as 1-(3,5-dichloro-2,6-dihydroxy-4-methoxyphenyl)-1-hexanone.

This work describes the discovery of a new type of biologically active chemical compound. It demonstrates the existence of a new class of effector molecule, whose lipid and water solubility may allow passage directly from cytoplasm to cytoplasm. Finally, knowledge of the structure and the availability of relatively large quantities of the synthetic substance and its analogues should now allow a full analysis of both the biosynthesis of DIF and its function in the pattern-forming processes during *Dictyostelium* development^{18–23}.

We are very grateful to Babu Dhokia, Colin Young and Angus Wilson for help in DIF production, to Chris Town and Gerry Weeks for discussion and to Julian Gross in whose laboratory, at ICRF Mill Hill, this project was started. We further thank Professor Wilhelm J. Richter for MS-MS, Ms Sue Johnson for NMR and Dr Bruce Chase for FT-IR analyses.

Received 13 April; accepted 29 June 1987.

1. Wolpert, L. *J. theor. Biol.* **25**, 1–47 (1969).
2. Gierer, A. & Meinhardt, H. *Kybernetik* **12**, 30–39 (1972).
3. Anderson, K. V. & Nüsslein-Volhard, C. *Nature* **311**, 223–227 (1984).
4. Fröhnhöfer, H. G. & Nüsslein-Volhard, C. *Nature* **324**, 120–125 (1986).
5. Wharton, K. A., Johansen, K. M., Xu, T. & Artavanis-Tsakonas, S. *Cell* **43**, 567–581 (1985).
6. Macdonald, P. M. & Struhl, G. *Nature* **324**, 537–545 (1986).

7. Greenwald, I. *Cell* **43**, 583-590 (1985).
8. Schaller, H. C. & Bodenmuller, H. *Proc. natn. Acad. Sci. U.S.A.* **78**, 7000-7004 (1981).
9. Maden, M. *Nature* **295**, 672-675 (1982).
10. Tickle, C., Alberts, B., Wolpert, L. & Lee, J. *Nature* **296**, 564-566 (1982).
11. Smith, J. C. *Development* **99**, 3-14 (1987).
12. Town, C. D., Gross, J. D. & Kay, R. R. *Nature* **262**, 717-719 (1976).
13. Town, C. D. & Stanford, E. *Proc. natn. Acad. Sci. U.S.A.* **76**, 308-312 (1979).
14. Brookman, J. J., Town, C. D., Jermyn, K. A. & Kay, R. R. *Dev. Biol.* **91**, 191-196 (1982).
15. Kay, R. R., Dhokia, B. & Jermyn, K. A. *Eur. J. Biochem.* **136**, 51-56 (1983).
16. Morris, H. R., Taylor, G. W., Piper, P. J. & Tappin, J. R. *Nature* **285**, 104-106 (1980).
17. Beloff-Chain, A., Morton, J., Dunmore, S., Taylor, G. W. & Morris, H. R. *Nature* **301**, 255-258 (1983).
18. Kay, R. R. & Jermyn, K. A. *Nature* **303**, 242-244 (1983).
19. Williams, J. G. *et al.* *Cell* **49**, 185-192 (1987).
20. Kopachik, W., Oohata, A., Dhokia, B., Brookman, J. J. & Kay, R. R. *Cell* **33**, 397-403 (1983).
21. Brookman, J. J., Jermyn, K. A. & Kay, R. R. *Development* **100**, 119-124 (1987).
22. Gross, J. D., Bradbury, J., Kay, R. R. & Peacey, M. J. *Nature* **303**, 244-245 (1983).
23. Sobolewski, A., Neave, N. & Weeks, G. *Differentiation* **25**, 93-100 (1983).

Currents through the fusion pore that forms during exocytosis of a secretory vesicle

L. J. Breckenridge & W. Almers

Department of Physiology and Biophysics, University of Washington
SJ-40, Seattle, Washington 98195, USA

Exocytosis, or the fusion of cytoplasmic vesicles with the cell membrane, occurs in nearly all eukaryotic cells, but its mechanism is not understood. Morphological^{1,2} and electrophysiological studies³⁻⁵ have suggested that membrane fusion begins with the formation of a 'fusion pore', a narrow channel across the closely adjacent membranes of vesicle and cell that forms the first connection of the vesicle lumen with the cell exterior and later dilates to allow release of vesicle contents. We used the patch clamp technique to study exocytosis of single giant secretory vesicles in mast cells of beige mice^{4,5}. The first opening of the fusion pore was found to generate a brief current transient, whose size and direction indicated an initial pore conductance of about 230 pS and a lumen-positive vesicle membrane potential. In time-resolved a.c. admittance measurements, the pore conductance was found to increase to much larger values within milliseconds, as if the pore dilated soon after opening. We conclude that the earliest fusion event may be the formation of a structure similar to an ion channel. Its conductance is of the same order of magnitude as that of a single gap junction channel, the only other known channel that spans two membranes.

The exocytosis of single secretory vesicles causes stepwise increases of the cell surface area, and this effect may be assayed by measuring the cell membrane capacitance, C_m (ref. 3). Figure 1a shows the results of measuring C_m in a single mast cell stimulated to secrete by cytosolic application of GTP γ S, a non-hydrolysable analogue of GTP. The value of C_m increased in steps, each reporting the fusion of a single vesicle with the cell membrane³⁻⁵, or, more precisely, the establishment of an electric connection between the lumen of an exocytosing vesicle and the cell exterior. The initial conductance of this connection, g_0 , may be measured as follows. Suppose that E_c and E_v , the potentials across cell and vesicle membranes respectively, differ initially (Fig. 1b). Once the connection is formed, E_v and E_c must equalize by changing the charge on the vesicle membrane capacitance (C_v , see Fig. 4c) by an amount Q :

$$Q = C_v(E_c - E_v) \quad (1)$$

where C_v is measured experimentally as ΔC , the final amplitude of the observed C_m step. The charge movement should appear as a transient current whose initial (peak) amplitude, I_0 , is related to g_0 by

$$g_0 = I_0 / (E_c - E_v) \quad (2)$$

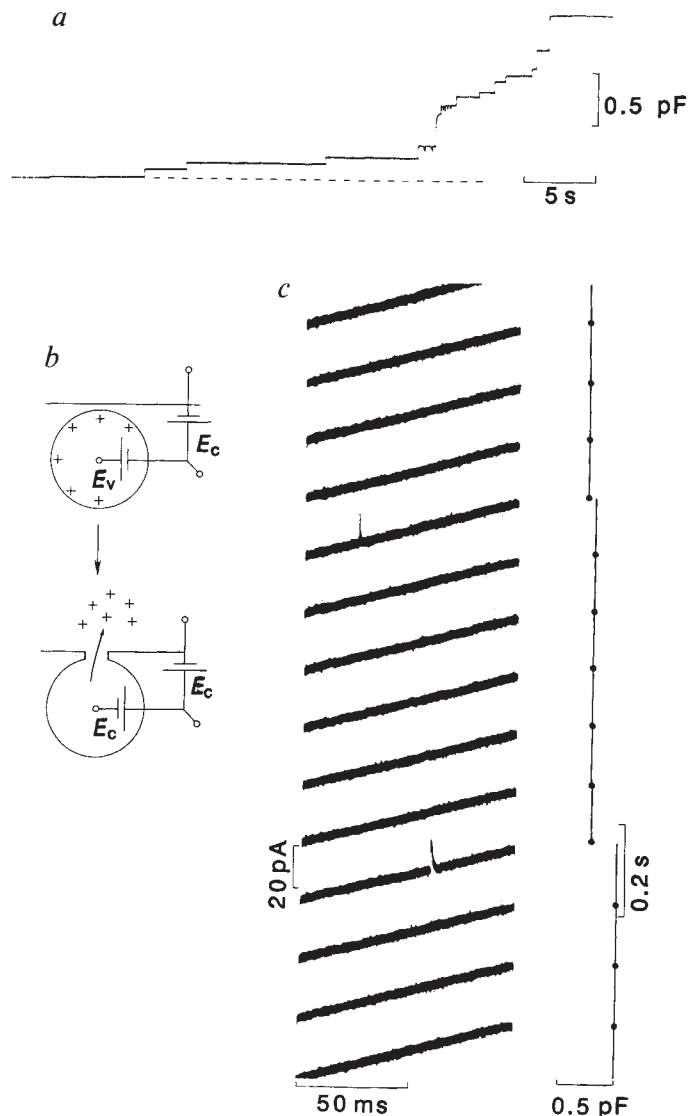


Fig. 1 *a*, Membrane capacitance plotted against time in a degranulating mast cell voltage-clamped at 0 mV. Each time a vesicle fuses with the cell membrane there is a step increase in C_m . *b*, Potentials across cell membrane (E_c) and vesicle membrane (E_v) defined as cytosol minus outside and cytosol minus lumen, respectively. *c*, Membrane current (thick traces) during a short portion of an experiment as in *a* except that the sinusoid used to measure C_m was applied only once every 112.5 ms for 12.5 ms at a time. Holding potential 10 mV. Vertical trace, mean C_m values during the successive 12.5-ms periods. Note C_m increases leftwards, time upwards; dots, time of measurement. Often, the increase in C_m due to exocytosis of a single vesicle continued over two or three 125-ms episodes (see also Fig. 4b); in such cases the final amplitude of the C_m step was used in the analysis.

Methods. Peritoneal mast cells were obtained from bg^i/bg^j mice⁵, a mutant with giant secretory vesicles. A firepolished micropipette was sealed against the cell, and the patch of membrane beneath the pipette tip ruptured by a pulse of suction ('whole-cell configuration'¹⁴) so that the cell could be voltage clamped and dialysed with the solution in the pipette (containing 155 mM potassium glutamate, 5 mM MgCl₂, 0.5 mM CaCl₂, 1 mM potassium EGTA, 4 mM potassium ITP, 10 mM sodium HEPES, pH 7.2, 20 μ M GTP γ S). External solution: 140 mM NaCl, 2 mM KCl, 2 mM CaCl₂, 5 mM MgCl₂, 5 mM glucose, 10 mM sodium HEPES buffer, pH 7.2. C_m was measured with a lock-in amplifier¹⁵ as the imaginary component of the inside-out admittance of the cell as in Fig. 4. Temperature 22-24°C. Measurements generally give $\pm$ s.e.m.

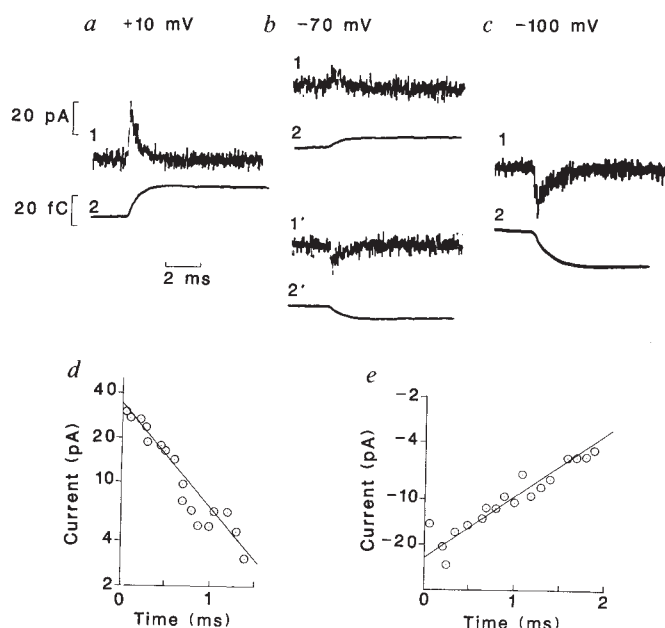


Fig. 2 Traces from three cells with holding potentials (E_c) indicated. Traces 1, 1', current transients recorded as in Fig. 1c. Traces 2, 2', time integrals of current traces from an analogue integrator. *d, e*, Semilogarithmic plots of *a1, c1* with time constants of 0.63 ms (*d*) and of 1.07 ms (*e*). Current filtered through a 10-kHz, 7-pole Bessel filter cascade. Steps of C_m (not shown) had final amplitudes of 170 fF (*a*), 242 fF (*b1, 2*), 219 fF (*b1', 2'*) and 294 fF (*c*). From E_c and equation (1), values for E_v in the four vesicles were: -118 mV (*a*), -100 mV (*b1, 2*), -34 mV (*b1', 2'*), -29 mV (*c*). Conductances in *a* and *c* were 251 pS and 285 pS from equation (2), and 272 pS and 275 pS calculated from the time constants. In eight cells held at potentials between 30 and 50 mV, 84% of all C_m steps were preceded by transients. Because current was recorded only 90% of the time we expect to have missed 10% of the transients, so the actual percentage of steps preceded by transients must have been $84/0.9 = 93\%$.

If g_0 does not change during the transient, current should decline exponentially with time constant τ , and the conductance can then also be calculated as the ratio $\Delta C/\tau$.

The heavy sloping lines in Fig. 1c show current traces measured during 14 successive episodes. At the end of each episode, a sinewave was applied briefly to the cell membrane to measure C_m . The vertical trace on the righthand side of Fig. 1c represents the time course of C_m over the period covered by the current traces on the left. While C_m remained constant, the traces showed silence, as expected from the paucity of ion channels in the mast cell membrane⁶. However, 2 of the 14 traces showed a brief, spike-like current transient, and after each, C_m was found to have increased. Figures 2a1-c1 show, on a faster time base, transients as in Fig. 1c recorded at different cell membrane holding potentials. The charge carried by the current transients is represented by their time integrals (Figs 2a2-c2).

The current transients are clearly associated with exocytosis of single vesicles, as represented by the C_m steps. They were never observed in quiescent cells. While cells were undergoing degranulation, an average of one out of 27 sweeps showed a transient. Out of 330 recorded current transients, all but one occurred precisely at the beginning of a C_m change that appeared stepwise on the time scale of Fig. 1a. Moreover, at the most positive cell membrane potentials investigated (30 to 50 mV), essentially all the 179 C_m steps recorded were preceded by a current transient. The chance of observing such consistent temporal correlation by accident is vanishingly small. If one out of 27 sweeps shows a transient, and one out of 27 is followed by the beginning of a C_m step, the chance of having both coincide 179 times in a row is 27^{-179} .

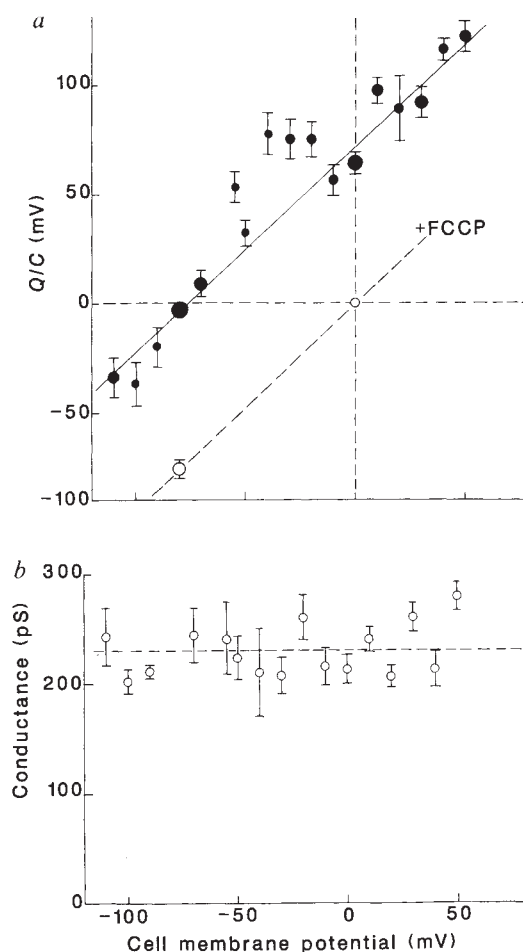


Fig. 3 *a*, Dependence of Q/C (equal to the difference between E_c and E_v) on E_c . Steps in C_m without visible transients (blanks) were assigned a charge of zero, except that a number of blanks equal to 10% of all steps was rejected because we expect to miss 10% of the transients. ●, Conditions as in Figs 1 and 2. The regression line (continuous) has a slope of 0.94 ± 0.4 and intercepts the abscissa at -76 ± 14 mV, the average potential of vesicles at the instant of fusion. Similar values were obtained by equation (1) when ATP replaced ITP in the pipette (not shown). The weight given to each point as well as the plotted size are in proportion to the number of cells included (between one and eight). ○, With 1 μ M FCCP added to the external solution between 15 and 45 min before the experiment; dotted line drawn by eye. Without FCCP, 81% of the steps in C_m at 0 mV cell membrane potential were accompanied by current transients, none of which were inward; at -80 mV, 13% of C_m steps had outward transients, 16% inward transients and the remainder none. In the presence of FCCP, none of the C_m steps at 0 mV gave a detectable transient, whereas 93% of the C_m steps at -80 mV gave large inward transients. *b*, Values of g_0 from equation (2) and from transients accompanying C_m steps with final amplitude 80–250 fF. Horizontal line at 230 pS. Symbols are averages of all values at a given holding potential.

The transients had the features expected if they represent the charging or discharging of a capacitive element. (1) They declined monotonically, often in an exponential fashion (Fig. 2). (2) At a given holding potential, the charge carried by the transient depended linearly on the final amplitude of the C_m step (not shown). (3) The ratio I_0/Q , for an exponentially declining transient equal to the time constant of decline, varied widely (range 0.15–2 ms) but was strongly correlated with the C_m -step amplitude (plot not shown, but see Fig. 1c). (4) When transients were recorded at different holding potentials (Fig. 2a–c), the ratio charge/ ΔC depended on the holding potential, the regression line having a slope close to unity (Fig. 3a).

Fig. 4 *a, b*, Imaginary (Im) and real (Re) parts of the a.c. admittances contributed by two exocytosing vesicles. Continuous traces plot sequences of discrete measurements, six of them separately indicated as dots in *a*. The first one or two pairs of Re and Im measurements at the beginning of each step were generally erratic from contamination by the current transient, whose occurrence they are taken to mark. *c*, Equivalent circuit of an exocytosing vesicle, with vesicle capacitance C_v and a variable conductance of the fusion pore, g . Its electrical admittance, $Y(\omega)$, based on the ratio of current over voltage, is^{4,16}:

$$Y(\omega) = \frac{(\omega C_v)^2/g}{(\omega C_v/g)^2 + 1} + j \frac{\omega C_v}{(\omega C_v/g)^2 + 1} \quad (3)$$

where the first and second terms represent Re and Im, respectively, in *a* and *b*. $\omega = 2\pi f$, $f = 800$ Hz and $j = \sqrt{-1}$. The circuit behaves as follows. When g is small, the sinusoidal voltage drops almost entirely across this element, hence the vesicle will appear electrically as a conductance, pass current nearly in phase with voltage and contribute mostly to Re. As g grows, so does Re, but the voltage will increasingly drop also across C_v ; the vesicle will begin to assume the electrical appearance of a capacitor and hence contribute to both Re and Im. When g is so large that the voltage drops entirely across C_v , the vesicle appears as a pure capacitor and contributes only to Im. Assuming C_v to be given by the final value of the Im change, g is obtained from either Re or Im:

$$g = \omega C_v / (\omega C_v / \text{Im} - 1)^{1/2} \quad (4)$$

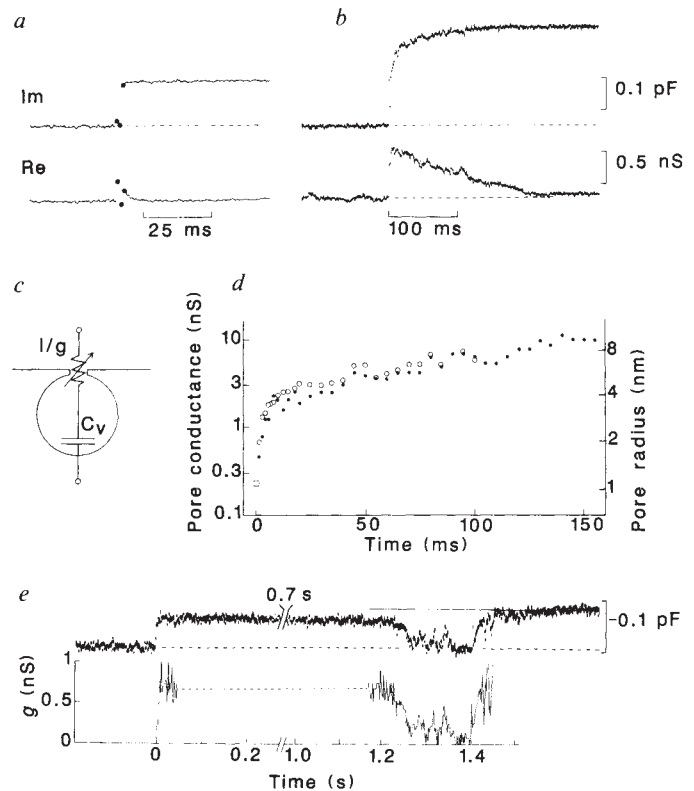
$$g = 2 \text{Re} / \{1 + n[1 - (2\text{Re}/\omega C_v)^2]^{1/2}\} \quad (5)$$

where $n = 1$ for $\text{Im} < \omega C_v/2$ and $n = -1$ for $\text{Im} \geq \omega C_v/2$. *d*, Plot of pore conductance against time for vesicle in *b*; circles calculated from (4) and dots from equation (5). The square represents the expected starting value of 230 pS. Right-hand ordinate, pore diameter calculated from equation (8-1) of ref. 12 for a resistivity of $100 \Omega \text{ cm}$ and a pore length of 15 nm. *e*, Im (top) and g (below) for a third vesicle showing 'flicker'; g calculated by equation (4) with C_v given by the final value of ΔC (continuous horizontal line). Values for $g > 1$ nS not plotted; dashed line at 700 pS.

Methods. A continuous 800-Hz sinewave of 52 mV peak amplitude was superimposed on the holding potential of the cell membrane (-50 mV in *a*, 0 mV in *b* and *e*). Magnitudes of the in-phase (Re) and 90° -out-of-phase (Im) components of membrane current were determined by a two-phase lock-in amplifier. Its outputs Re and Im were each sampled at 800 Hz; each sample represents the time average of Re or Im over one sinusoidal cycle (1.25 ms) as obtained by an appropriately phase-locked analogue integration over that period. Immediately after attaining whole-cell configuration, the sinusoidal current across the cell membrane was nulled out with a compensation circuit; this allowed recording at a higher gain and also determined the resistance of the pipette-cell junction and the initial value of C_m . On the basis of these two values and a calibration table, the phase of the lock-in amplifier was then adjusted to compensate for phase lags introduced by equipment and the pipette resistance.

At positive cell membrane potentials where transients were largest, the value for g_0 from equation (2) was 230 ± 83 pS (mean $\pm$ s.d., $n = 237$). Similar values were also found at other cell membrane potentials (Fig. 3*b*). The variability was large, with a suggestion of larger values for larger vesicles. From the rising phase of the transient, too rapid to be well resolved in Fig. 2*a1-c1*, it appears that the fusion pore can form in less than 0.1 ms. For seven vesicles with exponentially declining transients, g was calculated also from time constants of decline, obtained from semilogarithmic plots of current transients (Fig. 2*d* and *e*) or their time integrals. During the first 0.5–1 ms of decline, the time constants gave values for g that were 1.08 ± 0.05 times larger than g_0 by equation (2); the difference is small, indicating that g_0 did not change greatly in these cases. However, semilogarithmic plots of other transients showed an increase in slope with time, as if the conductance increased during the transient.

Membrane potentials of exocytosing vesicles at the instant of fusion were calculated by equation (1) and found to vary widely (range -11 to -160 mV, cytosol minus lumen). Even in the same cell, different vesicles had different potentials; this may be seen in Fig. 2*b* where two transients in opposite directions were recorded within 1 s of each other. In Fig. 3*a*, E_v has a mean value of -76 mV. X-ray microanalysis of mast cell granules and cytoplasm⁷ suggests that none of the major inorganic ions (Cl, K, Na) have a concentration gradient of the appropriate size or direction to account for a large lumen-positive potential across the vesicle membrane. If the vesicle membrane were selectively H^+ permeable, a lumen-negative potential would be



expected because the vesicle lumen is acidic compared to the cytoplasm⁸. The potential observed here could be the result of an electrogenic H-pump in a vesicle membrane of low ion permeability. A lumen-positive potential depending on the presence of ATP has been inferred also for isolated chromaffin granules from measurements with potentiometric dyes⁹.

The H^+ ionophore FCCP (carbonylcyanide-4 (trifluoromethoxy)-phenylhydrazine) is expected to increase the permeability of the vesicle membrane to H^+ and collapse or even reverse the potential across it. Indeed, mast cells incubated with FCCP (circles in Fig. 3*a*) had a vesicle membrane potential close to 0 mV. This large effect of FCCP is strong evidence that the current transients result from an intracellular membranous compartment, such as secretory vesicle. Because cytosolic and extracellular ion concentrations are controlled under our conditions, FCCP is not expected to change the reversal potential of any ion channel in the cell membrane. Interestingly, mast cells in FCCP readily degranulated. Hence the experiments in Fig. 3 also showed that exocytosis readily occurs over a wide range of potentials across both vesicle and cell membranes.

Evidently, fusion starts with an aqueous pore of about 230 pS mean conductance, similar to that of a single gap junction channel (160 pS, ref. 10) whose pore diameter is < 2 nm (ref. 11). In electron micrographs, even the smallest pores observed are more than ten times larger (50 nm in mast cells¹). The comparison suggests that the fusion pore inferred here dilates soon after it is formed. To test this possibility and measure the pore conductance after the transient had declined, a sinewave was applied to the cell membrane. The corresponding sinusoidal

current was analysed with a lock-in amplifier that provided a signal proportional to the component in phase with the cell membrane potential (real component, labelled 'Re' in Fig. 4a and b) and another 90° out of phase ('Im' or imaginary component). An increase in the cell membrane conductance (for example, by an ion channel) results in a proportional increase in Re, whereas an increase in C_m (due for example to a completely exocytosed vesicle) results in a proportional increase in Im. During the early stages of exocytosis, however, a secretory vesicle is best represented⁴ by a resistor (that of the fusion pore) in series with a capacitor (the vesicle membrane, see Fig. 4c), a combination expected to contribute both real and imaginary components. Theory predicts that, as g grows from zero to infinity, Im should grow to a final value (proportional to ΔC) whereas Re should rise and then decline to zero.

Figure 4 shows changes in Im and Re caused by exocytosis of a small vesicle (a) and a larger one (in b). As predicted, exocytosis caused no permanent increase in Re. Evidently, (1) the vesicle membranes contributed no significant conductance and hence contained few or no open ion channels, and (2) the resistance in series with the vesicle membrane ultimately became negligible, consistent with the formation of a wide pore. In Fig. 4a, Im abruptly increased to its final value, whereas Re rapidly declined to zero. In this fusion, the expected changes in Im and Re evidently were too fast to be well resolved; by equation (5), g must have increased to over 6.7 nS within 3.75 ms, the time of the earliest reliable Re measurement after electrical changes began in Fig. 4a. In Fig. 4b, the changes in Re and Im were slower and allowed us to reconstruct the time course of g (Fig. 4d). The pore conductance increased to over 500 pS in 2.5 ms, to 2,500 pS in 10 ms and then continued to grow more slowly. For the right-hand ordinate of Fig. 4d, g is converted into pore radius¹², assuming the conductance to be due to a single dilating pore with the length of a gap junction¹¹. After about 150 ms, the pore diameter approaches the range observed in electron micrographs¹. In 17 similar analyses, the value of g determined between 2.5 and 3.75 ms after the probable occurrence of the current transient had a median value of about 1 nS (range 170 pS to >10 nS). Thus the initially small (230 pS) conductance in series with the vesicle membrane usually increases rapidly during exocytosis, consistent with a rapid dilation of the fusion pore.

In Fig. 4e, Im increased abruptly, but later fluctuated, repeatedly returned to baseline and finally stabilized at a higher level. Such 'capacitance flicker' was previously taken as evidence of an early and reversible step in membrane fusion, representing the opening and closing of a fusion pore³ narrow enough to retard the escape of fluorescent dye⁵. By equation (4) (see Fig. 4, legend), the conductance in series with the flickering vesicle was about 700 pS, as if due to a pore of 4 nm diameter. The traces suggest that the fusion pore can dilate to 4 nm, persist in that state for a second and still close completely. Calculations show (ref. 12) that a small molecule dissolved within a 4 μ m giant vesicle would escape through such a pore only slowly (time constant 2 min), as observed⁵. A normal mast cell vesicle (0.8 μ m diameter¹³) with its much smaller volume, however, would discharge such a molecule with a time constant of only about 1 s, and hence could release much of its histamine without ever undergoing complete fusion.

In conclusion, exocytosis of a secretory vesicle begins with the sudden formation of a pore with a conductance of ~230 pS. The conductance rapidly increases, suggesting that the pore dilates. The variable and transient conductance of the initial pore suggests that it arises from a variable number of membrane protein subunits that aggregate loosely in a ring, forming a channel in their midst. The initial diameter of the channel is probably little larger than that of a gap junction channel.

We thank Drs E. Neher and J. Zimmerberg for helpful discussions, Drs P. Detwiler and B. Hille for reading our manuscript and R. Reinstatler for construction of electronic equipment.

Received 17 April; accepted 30 June 1987.

1. Chandler, D. E. & Heuser, J. E. *J. cell Biol.* **86**, 666-674 (1980).
2. Ornberg, R. L. & Reese, T. S. *J. cell Biol.* **90**, 40-54 (1981).
3. Fernandez, J. M., Neher, E. & Gomperts, B. D. *Nature* **312**, 453-455 (1984).
4. Zimmerberg, J., Curran, M., Cohen, F. S. & Brodwick, M. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1585-1589 (1987).
5. Breckenridge, L. J. & Almers, W. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1945-1949 (1987).
6. Lindau, M. & Fernandez, J. M. *J. gen. Physiol.* **88**, 349-368 (1986).
7. Kendall, M. D. & Warley, A. *J. cell Sci.* **83**, 77-87 (1986).
8. Johnson, R. G., Carty, S. E., Fingerhood, B. J. & Scarpa, A. *FEBS Lett.* **120**, 75-79 (1980).
9. Salama, G., Johnson, R. G. & Scarpa, A. *J. gen. Physiol.* **75**, 109-140 (1980).
10. Veenstra, R. D. & De Haan, R. L. *Science* **233**, 972-974 (1986).
11. Unwin, P. N. T. & Zampighi, G. *Nature* **283**, 545-549 (1980).
12. Hille, B. *Ion Channels of Excitable Membranes*, 186-188 (Sinauer Associates, Sunderland, Massachusetts, 1984).
13. Helander, H. F. & Bloom, G. D. *J. Microscopy* **100**, 315-321 (1974).
14. Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).
15. Neher, E. & Marty, A. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6712-6716 (1982).
16. Falk, G. & Fatt, P. *Proc. R. Soc.* **160**, 69-123 (1964).

Production and auto-induction of transforming growth factor- α in human keratinocytes

Robert J. Coffey Jr, Rik Derynck*, Josiah N. Wilcox*, Timothy S. Bringman*, A. Scott Goustin†, Harold L. Moses & Mark R. Pittelkow‡

Departments of Cell Biology and Medicine, Vanderbilt University School of Medicine, Nashville, Tennessee 37232, USA

* Department of Molecular Biology, Genentech Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080, USA

† Department of Cell Biology and ‡ Department of Dermatology, Mayo Clinic/Foundation, Rochester, Minnesota 55905, USA

Transforming growth factor- α (TGF- α) is a polypeptide which is structurally related to epidermal growth factor (EGF)^{1,2} and binds to the EGF receptor^{3,4}. TGF- α synthesis occurs in a variety of neoplastic cells⁵ and during early fetal development^{6,7} but has not been reported in normal cells of the adult organisms. TGF- α has therefore been regarded as an embryonic growth factor which is inappropriately expressed during neoplasia. Here we report that primary cultures of normal human keratinocytes synthesize TGF- α . Furthermore, we show that addition of EGF or TGF- α to these cultures induces TGF- α gene expression, suggesting that a mechanism of auto-induction exists. Analysis of normal skin biopsies using *in situ* hybridization and immunohistochemistry demonstrates the *in vivo* presence of TGF- α messenger RNA and protein in the stratified epidermis.

TGF- α was first discovered in the medium of retrovirally transformed fibroblasts⁸, but it has since been demonstrated that a large variety of neoplastic cells and tumours of non-haematopoietic origin display TGF- α expression⁵. In contrast, it has not been demonstrated that normal cells synthesize TGF- α *in vivo*, although it is produced by bovine anterior pituitary cells in culture⁹. The TGF- α gene is also expressed during early fetal development⁶. These observations have led to the proposal that TGF- α could have a function in the transformation process by an autocrine feedback effect on the cells which synthesize TGF- α ¹⁰, a hypothesis supported by the observation that expression of TGF- α in immortalized fibroblasts can result in malignant transformation¹¹. The expression of TGF- α by squamous carcinoma cells⁵ of epithelial origin led us to investigate the synthesis of TGF- α by normal human keratinocytes *in vitro* and *in vivo*.

Human keratinocytes and melanocytes were derived from newborn human foreskins. Keratinocytes were cultured in a serum-free medium containing insulin, EGF and bovine pituitary extract (BPE) (complete medium)¹³; melanocytes were

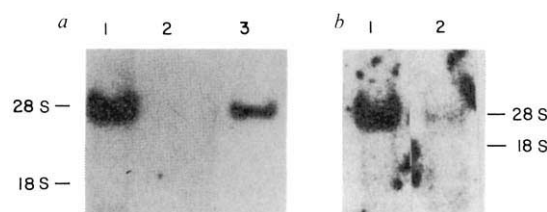


Fig. 1 Detection of TGF- α mRNA in primary cultures of skin keratinocytes and in skin tissue. *a*, Northern blot analysis of poly(A) RNA (2 μ g) from cultures of human keratinocytes (lane 1), human melanocytes (lane 2) and from a colon carcinoma cell line SW620 (lane 3). *b*, Northern blot analysis of poly(A) RNA from primary culture of foreskin-derived human keratinocytes (lane 1, 2 μ g) and neonatal foreskin tissue (lane 2, 9 μ g).

Methods. Neonatal human proliferative keratinocytes were cultured in serum-free medium containing insulin, EGF and BPE as described previously¹³. Melanocytes were cultured in this medium with the addition of TPA (10 ng ml⁻¹)¹⁴. RNA was extracted from the cultures by the method of Schwab *et al.*²⁷ and prepared from frozen foreskin tissue which was pulverized in a Waring blender cup containing liquid nitrogen. The tissue powder was suspended in NETS buffer (0.1 M NaCl, 10 mM Tris-HCl, pH 7.4, 1 mM EDTA, 0.2% SDS) in the presence of 10 mM vanadyl ribonucleoside complex (BRL) and extracted in phenol. DNA was removed by repeated precipitation of the RNA from 2.25 M sodium acetate, pH 6.0. Oligo(dT)-selected RNA was separated by electrophoresis in a 1.2% agarose-formaldehyde gel²⁸ and Northern blotting performed as described²⁹. The human TGF- α complementary DNA probe corresponded to the cDNA insert of plasmid pTGF-C1 (ref. 2) and contains the complete coding sequence. The probe was subcloned into plasmid SP65 and labelled with [α -³²P]UTP as described³⁰. All hybridizations and post-hybridization washes were performed at 65 °C as described¹⁵.

cultured in this medium with the addition of phorbol 12-myristate 13-acetate (TPA; 10 ng ml⁻¹)¹³. Polyadenylated RNA was isolated from rapidly growing keratinocyte and melanocyte cultures and analysed by Northern hybridization for the presence of TGF- α mRNA (Fig. 1*a*). The keratinocyte sample (lane 1) contained a 4.5–4.8 kilobase (kb) TGF- α mRNA, which is consistent in transcript size with TGF- α mRNA from SW620 cells (lane 3) which express TGF- α ^{14,15}. The melanocyte sample (lane 2), however, did not contain any detectable TGF- α mRNA. Keratinocyte cultures derived from adult skin also showed this 4.5–4.8-kb signal (data not shown). An additional 2.4-kb mRNA appeared in some keratinocyte samples (data not shown). The nature of this transcript is unknown but it may be a result of premature polyadenylation as several potential polyadenylation signals are present in the 3'-untranslated sequence of the full-size TGF- α mRNA¹⁶. The levels of TGF- α secreted into the medium were evaluated: cultures of keratinocytes grown to confluence in complete medium showed a time-dependent increase in TGF- α secreted into the medium, 73 pg ml⁻¹ of TGF- α being detected in 24 h (Fig. 2). TGF- α protein was measured by enzyme-linked immunosorbent assay (ELISA) which is specific for TGF- α and does not recognize EGF¹⁷. TGF- α was not found in serum-free medium, in medium conditioned by melanocytes, or in the growth factor supplements.

Experiments in which insulin, EGF or BPE were individually deleted from the serum-free medium showed TGF- α expression was enhanced only by EGF and not by the other two growth factor supplements (data not shown). Keratinocytes cultured for 48 h in basal medium lacking the three growth-factor supplements, followed by addition of EGF to fresh basal medium without insulin or BPE, showed a time- and dose-dependent induction of TGF- α secretion into the medium (Fig. 3). Enhancement of TGF- α gene expression by EGF was monitored by mRNA hybridization. The low levels of TGF- α mRNA increased in a dose-dependent manner with the addition of EGF to the medium (Fig. 3*a*) and showed a correlation with the

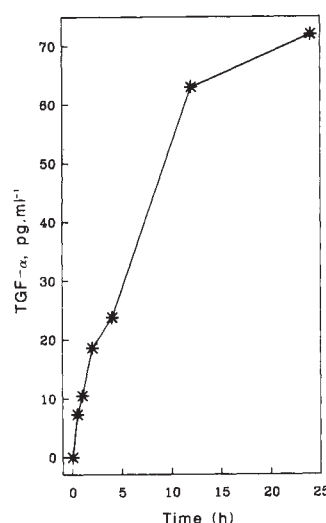


Fig. 2 Time-dependent secretion of TGF- α . Keratinocytes were cultured as described in legend to Fig. 1. The medium was removed from subconfluent cultures and replaced by fresh complete medium at time 0. Amounts of secreted TGF- α were measured at the indicated times by a TGF- α specific double sandwich ELISA¹⁷.

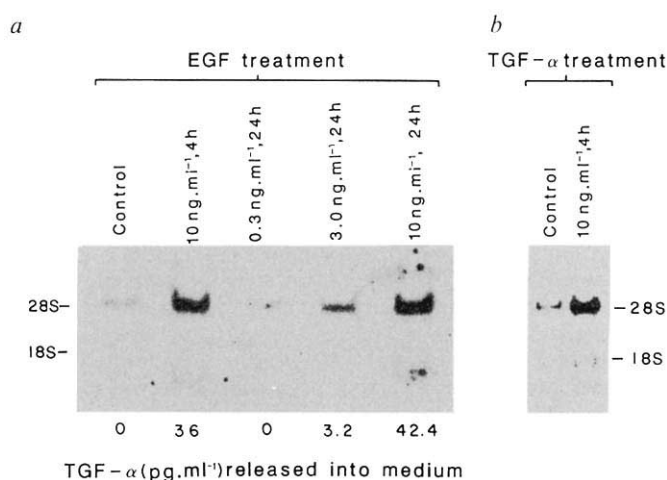


Fig. 3 Exogenous EGF and TGF- α enhance the levels of TGF- α mRNA in cultured keratinocytes. *a*, Effect of different doses of exogenous EGF on the amount of TGF- α mRNA. The levels of secreted TGF- α were measured in the medium by a double sandwich ELISA¹⁷; the values are indicated in pg ml⁻¹ below each Northern hybridization lane. *b*, Effect of 10 ng ml⁻¹ of exogenous TGF- α on the amount of TGF- α mRNA. The levels of newly synthesized TGF- α could not be measured because of interference by the exogenous TGF- α .

Methods. Keratinocytes were cultured and deprived at 60% confluency of insulin, EGF and BPE for 48 h. This medium was removed, fresh basal medium plus EGF or TGF- α added at the indicated concentrations and the medium and cells collected. Poly(A) RNA was prepared and Northern hybridizations performed as indicated in Fig 1 legend.

amount of TGF- α in the medium. Basal medium (without EGF) caused no stimulation of TGF- α mRNA or protein production.

As TGF- α and EGF are structurally related and interact with the same receptor, we examined whether addition of exogenous TGF- α would induce TGF- α expression in keratinocytes. Like EGF, TGF- α was able to enhance levels of TGF- α mRNA (Fig. 3*b*). Newly-synthesized TGF- α could not be measured owing to interference by exogenous TGF- α . In a separate experiment, RNA collected from keratinocyte cultures treated with purified heparin binding growth factor-2 (HBGF-2, or basic

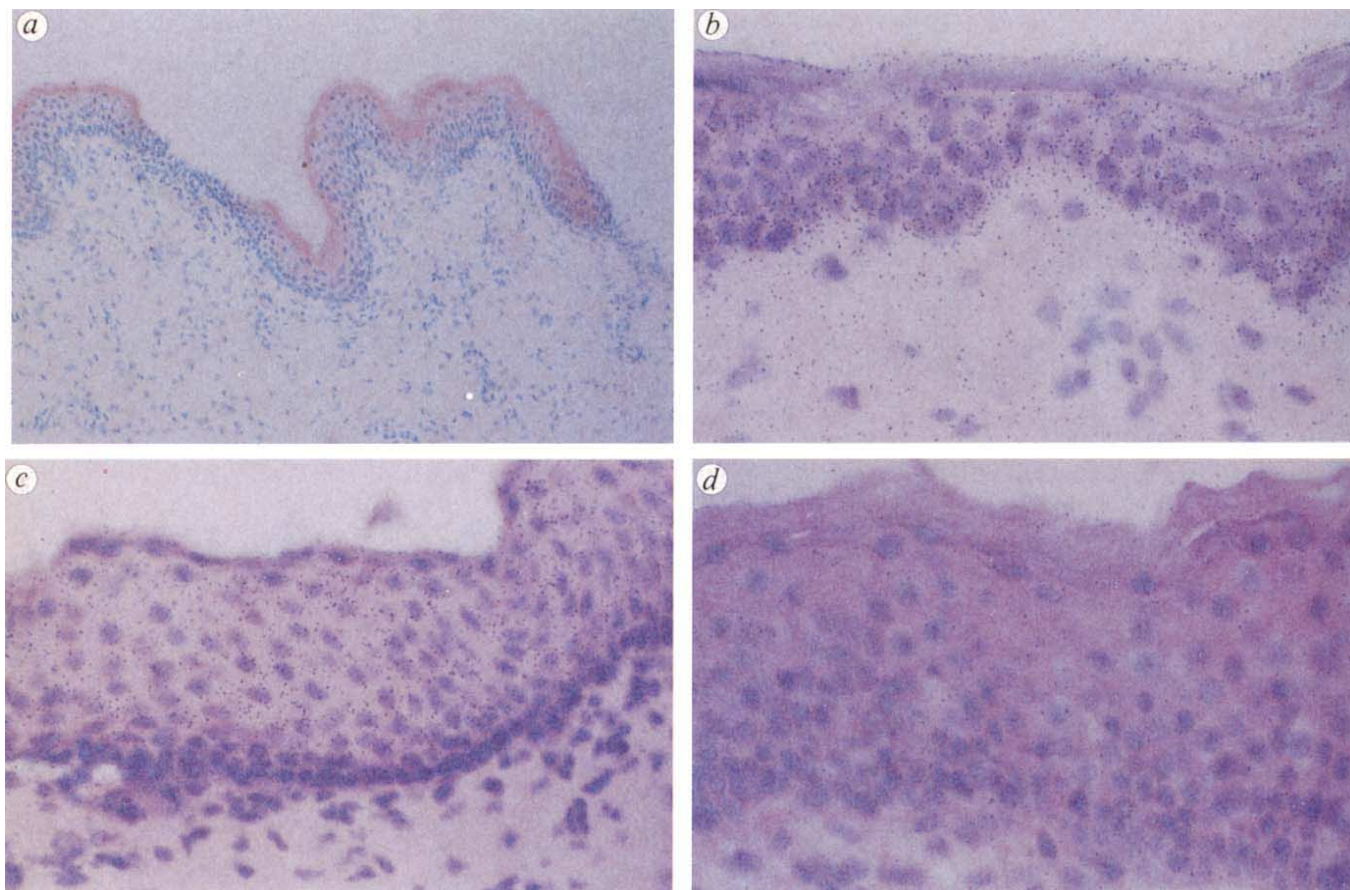


Fig. 4 Localization of TGF- α expression in histological sections of skin tissue by *in situ* hybridization and immunohistochemistry. *a*, TGF- α protein was detected by immunocytochemistry throughout the epidermis in neonatal foreskin. TGF- α mRNA was localized by *in situ* hybridization in normal adult breast skin (*b*) and neonatal foreskin (*c*). *d*, Control hybridization of a TGF- β cDNA to neonatal foreskin.

Methods. Immunohistochemistry was performed as previously described³¹ on Carnoy's fixed tissue using the Vectastain ABC alkaline phosphatase system (Vector) and a TGF- α monoclonal antibody (TGF- α 3 which was raised against human TGF- α 32-50 peptide AP3 as described previously¹⁷) at a concentration of $10 \mu\text{g ml}^{-1}$. The final reaction product was stained with the alkaline phosphatase substrate kit I to give a red final reaction product. Similar results were seen with polyclonal antisera 34D raised against purified human TGF- α ¹⁷, or antisera 34AP3 raised against human TGF- α 32-50 (ref. 17). No specific staining was seen using a control monoclonal antibody (human growth hormone) or when the primary antibody was not included in the reaction. Competition studies with recombinant TGF- α protein or synthetic peptide AP3 eliminated staining. *In situ* hybridization was performed using a human ³⁵S-labelled TGF- α riboprobe, described in Fig. 1 legend. The probe was labelled by *in vitro* transcription³⁰ in the presence of 500 μM each of CTP, GTP, UTP and 12 μM ³⁵S-ATP (Amersham: specific activity 650 Ci mmol⁻¹), 40 units RNasin in 2 mM spermidine, 10 mM dithiothreitol (DTT), 10 mM NaCl, 40 mM Tris, pH 7.5, and 15 units SP6 RNA polymerase. Hybridization was carried out overnight at 50 °C as previously described³¹. Post-hybridization treatments include incubation of the sections for 30 min at room temperature with 20 $\mu\text{g ml}^{-1}$ RNase A (Sigma) followed by washing in 0.2 $\times$ SSC at 42 °C for 2 h. Slides were exposed for 3 weeks. All TGF- α hybridizations were controlled by parallel hybridizations on serial sections to a labelled TGF- β riboprobe. In no case was specific labelling seen with the TGF- β probes (panel *d*). Magnification: $\times 365$.

fibroblast growth factor) (10 ng ml^{-1}) for 24 h showed TGF- α expression comparable to cultures deprived of growth factors for 48 h. Cultures treated with EGF (10 ng ml^{-1}) for 24 h showed the expected increase in TGF- α expression (data not shown). HBGF-2 at this dose is known to be mitogenic for cultures of human keratinocytes in clonal growth assays (G. D. Shipley and M.R.P., in preparation).

The expression of TGF- α by keratinocytes was also evaluated *in vivo*. Northern hybridization revealed that TGF- α mRNA was present in tissue samples of human foreskins (Fig. 1*b*) but at concentrations lower than in culture because whole neonatal foreskin contains connective tissue, non-epithelial cells (melanocytes) and dermal elements, in addition to the epidermal keratinocytes. The sites of TGF- α synthesis were localized by *in situ* hybridization (Fig. 4) of TGF- α mRNA in the epithelial cell layers of the skin, although no specific assignment of the cells could be made. Immunohistochemistry revealed that TGF- α immunoreactive material was present throughout the stratified epidermis (Fig. 4).

Our results demonstrate that TGF- α expression occurs *in vitro* and *in vivo* in a normal neonatal and adult cell type, the epidermal keratinocyte. We present evidence that the TGF- α expression in these cells is enhanced by exogenous EGF or TGF- α . The presence of EGF receptors on the surface of keratinocytes in culture^{18,19} and *in vivo*²⁰ suggests that the TGF- α synthesized by the keratinocytes can exert an autocrine or paracrine stimulation as a normal event, without associated malignant transformation. EGF and TGF- α have been found to induce proliferation of epidermal keratinocytes in culture (ref. 18 and Y. Barrandon and H. Green, personal communication). Therefore, endogenous TGF- α production could contribute to the *in vivo* proliferation of the normal keratinocytes of the skin.

Previous studies have shown that a growth factor can influence the expression of growth regulatory peptides, such as the induction of platelet-derived growth factor (PDGF)-like molecules by TGF- β ²¹. But the induction of TGF- α expression by EGF or TGF- α , which are structurally related and interact with the

same receptor, represents the first documented case of a growth factor inducing its own expression. It is as yet not known whether this effect is due to enhanced transcription or to mRNA stabilization.

This novel mechanism of auto-regulation of cell proliferation could be responsible for amplification of the growth factor response. Proper modulation of growth control might also require the regulated expression of growth inhibitory molecules by these same cells or by other types of cells which can exert their growth inhibitory activities in a paracrine way, as suggested after demonstration of expression of β -interferon-like molecules in BALB/c-3T3 cells by PDGF²². TGF- β , which is structurally unrelated to TGF- α , exerts a growth inhibitory activity on the keratinocytes²³ and could regulate growth of keratinocytes *in vivo*.

The induction of TGF- α expression by either EGF or TGF- α could be relevant in wound healing. It has been shown that EGF-like peptides which bind to EGF receptors are present in platelets²⁴ and their release at the site of wound healing could induce an enhanced TGF- α synthesis in keratinocytes to increase their rate of proliferation and augment the wound-healing process. The possible existence *in vivo* of a mechanism for auto-induction of growth factor expression may also have relevance in other pathological situations. Lack of adequate expression or effects of growth inhibitors could result in an abnormally high TGF- α expression which could then contribute to an increased rate of proliferation of the keratinocytes, as seen in psoriasis²⁵. High levels of EGF receptors are expressed in the keratinocytes throughout the psoriatic skin, but expression in normal skin predominates in the basal cell layer^{20,26}. Imbalance between amplification of the TGF- α expression and an inadequate mechanism of growth inhibition may be relevant in the malignant conversion of keratinocytes into squamous carcinomas, which consistently show relatively high levels of TGF- α expression⁷.

We thank N. N. Swanson, R. Graves-Deal, J. Wolfshohl and K. Smith for technical assistance, T. Spelsberg and P. Schein for support, and V. Murphy for typing the manuscript. This work was supported in part by grants from the National Cancer Institute and the J. W. Kieckhefer Foundation.

Received 29 January; accepted 1 July 1987.

- Marquardt, H., Hunkapiller, M. W., Hood, L. E. & Todaro, G. J. *Science* **223**, 1079-1082 (1984).
- Derynck, R., Roberts, A. B., Winkler, M. E., Chen, E.-Y. & Goeddel, D. V. *Cell* **38**, 287-297 (1984).
- Todaro, G. J., Fryling, C. & De Larco, J. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5258-5262 (1980).
- Massague, J. *J. biol. Chem.* **258**, 13614-13620 (1983).
- Derynck, R. *et al. Cancer Res.* **47**, 707-712 (1987).
- Lee, D. C., Rochford, R. M., Todaro, G. J. & Villareal, L. P. *Molec. cell. Biol.* **5**, 3644-3646 (1985).
- Twardzik, D. R. *Cancer Res.* **45**, 5413-5416 (1985).
- De Larco, J. E. & Todaro, G. J. *Proc. natn. Acad. Sci. U.S.A.* **75**, 4001-4005 (1978).
- Samsoondar, J., Kobrin, M. S. & Kudlow, J. E. *J. biol. Chem.* **261**, 14408-14413 (1986).
- Sporn, M. B. & Todaro, G. J. *New Engl. J. Med.* **303**, 878-880 (1980).
- Rosenthal, A., Lindquist, P. B., Bringman, T. S., Goeddel, D. V. & Derynck, R. *Cell* **46**, 301-309 (1986).
- Wille, J. J., Pittelkow, M. R., Shipley, G. D. & Scott, R. E. *J. cell. Physiol.* **121**, 31-44 (1984).
- Eisinger, M. & Marko, O. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2018-2022 (1982).
- Coffey, R. J., Shipley, G. D. & Moses, H. L. *Cancer Res.* **46**, 1164-1169 (1986).
- Coffey, R. J. *et al. Cancer Res.* (in the press).
- Derynck, R. in *Oncogenes, Genes and Growth Factors* (ed. Guroff, G.) 133-163 (Wiley, New York, 1987).
- Bringman, T. S., Lindquist, P. B. & Derynck, R. *Cell* **48**, 429-440 (1987).
- Rheinwald, J. G. & Green, H. *Nature* **265**, 421-424 (1977).
- O'Keefe, E., Battin, T. & Payne, R. *J. invest. Derm.* **78**, 482-487 (1982).
- Nanney, L. B., McKanna, J. A., Stoscheck, C. M., Carpenter, G. & King, L. E. *J. invest. Derm.* **82**, 165-169 (1984).
- Leof, E. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2453-2457 (1986).
- Zullo, J. N., Cochran, B. N., Huang, A. S. & Stiles, C. D. *Cell* **43**, 793-800 (1985).
- Shipley, G. P., Pittelkow, M. R., Wille, J. J., Scott, R. E. & Moses, H. L. *Cancer Res.* **46**, 2068-2071 (1986).
- Oka, Y. & Orth, D. N. *J. clin. Invest.* **72**, 249-259 (1983).
- Weinstein, G. D., McCullough, J. L. & Ross, P. A. *J. invest. Derm.* **85**, 579-583 (1985).
- Nanney, L. B., Stoscheck, C. M., Magid, M. & King, L. E. *J. invest. Derm.* **86**, 260-265 (1986).
- Schwab, M., Alitalo, K., Varmus, H. E. & Bishop, J. M. *Nature* **303**, 497-501 (1983).
- Donner, P. R., Kawasaki, E. S., Yu, L.-Y. & Bancroft, F. C. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2230-2234 (1981).
- Thomas, P. S. *Proc. natn. Acad. Sci. U.S.A.* **77**, 5201-5205 (1980).
- Melton, D. A. *et al. Nucleic Acids Res.* **12**, 7035-7056 (1984).
- Mason, A. J. *et al. Science* **234**, 1372-1378 (1986).

Requirement for intrinsic protein tyrosine kinase in the immediate and late actions of the EGF receptor

William S. Chen*†, Cheri S. Lazar*, Martin Poenie‡, Roger Y. Tsien‡, Gordon N. Gill* & Michael G. Rosenfeld*§

§ Howard Hughes Medical Institute, * Eukaryotic Regulatory Biology Program, School of Medicine M-013, † Chemistry Department, University of California San Diego, La Jolla, California 92093, USA
‡ Department of Physiology/Anatomy, University of California Berkeley, Berkeley, California 94720, USA

The epidermal growth factor (EGF) receptor is a transmembrane glycoprotein of relative molecular mass 170,000 with intrinsic ligand-dependent protein tyrosine kinase activity¹⁻⁶. Binding of EGF to its receptor activates a number of immediate biochemical processes, such as alterations of intracellular free calcium, pH, and increased transcription of several responsive genes, which usually culminate many hours later in DNA replication and cell division⁷⁻¹⁴. Abolishing the tyrosine kinase activity of three related oncogenes, *v-src*, *v-mos*, and *v-fps*, eliminates their capacity to transform cells¹⁵⁻¹⁷. Several reports have suggested that specific aspects of EGF receptor function are independent of the intrinsic tyrosine kinase activity^{18,19}; however, these studies used an antibody against EGF receptor which failed to activate phosphorylation of exogenous substrates¹⁸ and an insertional mutation in the EGF receptor tyrosine kinase domain which had not been shown to abolish protein kinase activity in cells¹⁹. Because many transmembrane receptors interact with intrinsic membrane proteins to activate second messenger systems, it is important to resolve experimentally whether mechanisms, in addition to activation of the intrinsic tyrosine kinase activity, mediate some EGF actions. From functional analyses of an EGF receptor containing a single amino-acid mutation at a site required for phosphate transfer from ATP, we conclude that the tyrosine kinase activity of the EGF receptor is essential for the diverse biochemical effects of EGF, including rapid alterations in intracellular calcium, activation of gene transcription, receptor down-regulation and the ultimate stimulatory effects on cell proliferation.

To test the functional consequences of abolishing tyrosine kinase activity with minimal structural alterations in the EGF receptor, the lysine at position 721 was mutated to a methionine. This amino acid, conserved amongst all the protein kinase family members, is thought to provide a salt bridge to β/γ phosphates of ATP in that binding site²⁰. The methionine for lysine substitution was selected because, as these two residues have similar bond lengths and angles, it should optimally preserve the overall structure of the EGF receptor protein. Both lysine 721 (K⁷²¹) and mutated methionine 721 (M⁷²¹) EGF receptor fusion genes contain an alanine at amino acid 654 which eliminates heterologous regulatory effects on the EGF receptor mediated by protein kinase C (ref. 21). A mutant dihydrofolate reductase gene²² provided a dominant selectable marker. Two cell lines which express no endogenous EGF receptor messenger RNA or protein were used for all studies—a mouse L cell line (B82)²¹ and Chinese hamster ovary (CHO) cells. Permanently transfected B82 and CHO cell lines used in these experiments expressed 350,000-450,000 and 25,000-35,000 human EGF receptors per cell, respectively. On Scatchard analysis of iodinated EGF binding²³, cells expressing functional K⁷²¹ and M⁷²¹ receptors had similar K_D for EGF binding (0.35-0.75 nM). Analysis of receptor protein by gel electrophoresis and Western blot analysis revealed the product of both constructs to be the expected holoreceptor glycoprotein of relative molecular mass (M_r) 170,000 (170 K) (data not shown). The mutation does not, therefore, affect synthesis or insertion of mature EGF-binding

Fig. 1 *a*, Autophosphorylation of the EGF receptor. A synthetic oligonucleotide (21-mer) was used to alter a single nucleotide (A to T) to generate a lysine (K) to methionine (M) substitution in amino acid 721 of an EGF receptor transcription unit under control of the simian virus 40 (SV40) early promoter enhancer¹⁹. A mutant dihydrofolate reductase (DHFR) gene provided the dominant selectable marker²². Permanent transfectants of B82 or CHO cells were selected as previously described¹⁹. Membranes (80 µg) from parental, M⁷²¹ and K⁷²¹ B82 cell lines and A431 cells were incubated in the presence of [γ -³²P]ATP and Mn²⁺, and analysed using SDS polyacrylamide gel electrophoresis (PAGE) and autoradiography³⁵. Lanes show phosphorylation in the absence (-) or presence (+) of 100 nM EGF using membranes from B82 cells expressing the K⁷²¹ or M⁷²¹ EGF receptor, or from untransfected parental cells (null). The migration of the 170K EGF holoreceptor in membranes of A431 cells (A431) is shown. *b*, Phosphorylation of angiotensin II by membranes (50 µg) from M⁷²¹, K⁷²¹ and null B82 cell lines. *c*, Effects of EGF on growth in K⁷²¹ (□, ■), M⁷²¹ (△, ▲), or null CHO cells (○, ●). Filled symbols, EGF treated cultures; open symbols, no EGF added. Population doubling times for each curve are: ○, 60 h; ●, 63 h; □, 66 h; ■, 36 h; △, 57 h; ▲, 66 h. *d*, Effects of EGF on expression of fusion genes containing the *c-fos* gene enhancer regulatory element. A plasmid containing two copies of -357 to 256 of the *c-fos*^H gene²⁴ fused to the luciferase complementary DNA²⁵ was transfected into B82 or CHO cell lines expressing the K⁷²¹ or M⁷²¹ EGF receptor as previously described¹⁹ using 5 µg plasmid per plate; 24 h later EGF (10 nM) was added and cells were harvested 24 h later. Luciferase activity was quantitated using described protocols²⁵; mock transfection background gave an identical value as assay blank. Results are the mean $\pm$ s.d. of triplicate determinations. Comparable results were obtained in both transfected B82 and CHO cell lines. Proof of responsiveness of M⁷²¹ cells was obtained by demonstrating that serum resulted in increased luciferase gene expression.

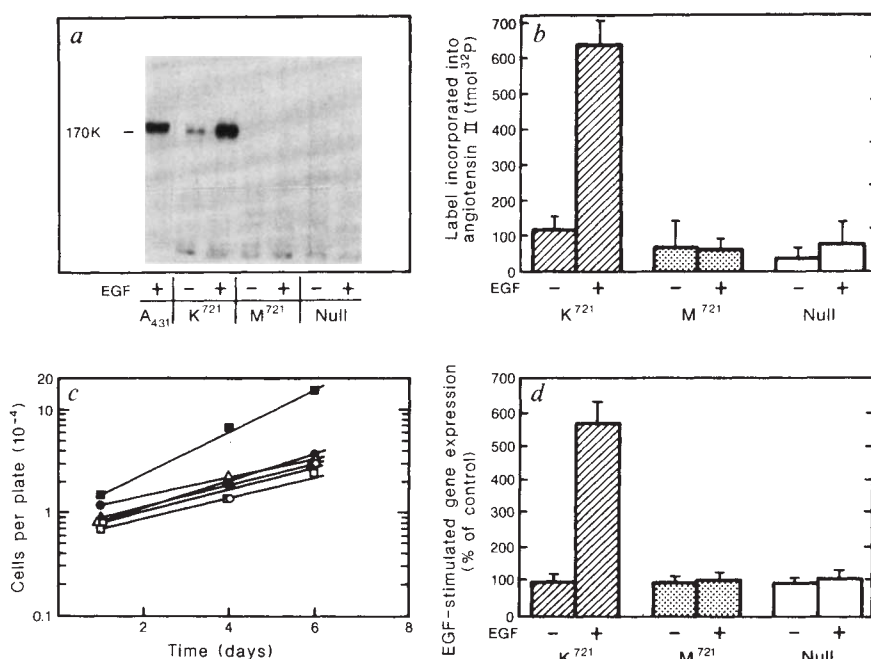
Methods. In *a*, autophosphorylation of EGF receptors in transfected cells was assayed using 80 µg membrane protein incubated 15 min at 22 °C in the absence or presence of 100 nM EGF in buffer containing 20 mM HEPES (pH 7.4), 0.6 mM EGTA, 4 mM 2-mercaptoethanol, and 12 µg ml⁻¹ leupeptin. Self-phosphorylation was then assayed at 4 °C by adjusting the buffer to 60 µM Na₃VO₄, 6 mM MnCl₂, and 3 µM [γ -³²P]ATP (15 mCi µmol⁻¹). Reactions were incubated for 5 min at 4 °C and analysed using 7.5% SDS-PAGE and autoradiography. Similar results were obtained in three independent experiments of similar design. In *b*, for phosphorylation of angiotensin II by B82 membrane preparations, membranes were incubated with 300 nM EGF for 15 min at 22 °C, and their ability to phosphorylate 2 mM angiotensin II quantitated by incubation in kinase reaction mixtures for 2 min at 4 °C. Reactions were terminated by addition of trichloroacetic acid to 6% and ³²P incorporation into angiotensin II was quantitated using phosphocellulose adsorption³⁶. Results are the average $\pm$ s.e.m. of triplicate determinations. In *c*, Cells were replated in serum-free medium containing 200 µg ml⁻¹ bovine serum albumin (BSA) grown in the presence or absence of EGF; and counted as previously described²¹. In *d*, The *fos* fusion genes were constructed by ligating two copies of synthetic oligonucleotides corresponding to nucleotides -357 to -276 of the *c-fos* gene as a *Bam*I-*Bgl*III fragment 5' of a TK(-200 to +70)-CAT fusion gene¹⁴. The fusion gene which contained *fos* enhancer, TK promoter and luciferase gene was constructed using pSV0 containing the LA5' luciferase gene²⁵, inserting the *fos* and TK sequences into the *Hind*III and *Nde*I sites.

receptors into the cell. Membranes prepared from untransfected B82 or CHO cells showed no EGF-dependent autophosphorylation or phosphorylation of exogenous substrates, consistent with their possessing no EGF receptors. Cell membranes from B82 cells expressing the K⁷²¹ EGF receptor exhibited a basal level of self-autophosphorylation which was markedly stimulated by EGF treatment (Fig. 1*a*). In contrast, membranes from B82 cells expressing the M⁷²¹ EGF receptor exhibited neither basal nor EGF-stimulated protein tyrosine kinase activity (Fig. 1*a*). Similarly, the K⁷²¹ EGF receptor exhibited EGF-dependent stimulation of phosphorylation of an exogenous substrate (angiotensin II) whereas the M⁷²¹ receptors showed no detectable kinase activity using exogenous substrates (Fig. 1*b*). These data indicate that the single substitution of methionine for lysine at amino-acid position 721 of the EGF receptor abolishes the intrinsic protein tyrosine kinase activity of the EGF receptor protein.

To investigate the effects of the mutant EGF receptor on growth or gene expression, transfected B82 or CHO cells expressing the EGF receptors were put in serum-free medium to minimize any growth effects caused by other mitogenic agents. Parental CHO cells and those expressing the M⁷²¹ EGF receptor gave no response to EGF (Fig. 1*c*) at concentrations from 0.01 to 10 nM. In contrast, transfected cell lines expressing the K⁷²¹ EGF receptor showed EGF-dependent growth, with an ED₅₀ for EGF-stimulated growth of 0.02 nM (data not shown), with

EGF decreasing population doubling time from 66 to 36 h (Fig. 1*c*). Similar results were observed in B82 cells. These results indicate that the tyrosine kinase activity of the EGF receptor is required for effects on cell proliferation.

Several biochemical events occur in responsive cells at the time of, or soon after, exposure to EGF. Binding of EGF to its cognate receptor rapidly stimulates transcription of a subset of genes, including proto-oncogenes such as *c-fos* and *c-myc*, and accumulation of their mRNAs^{9-14,24}. The region of the *c-fos* gene which transfers regulation by serum and growth factors to heterologous transcription units has been localized to a short 5' flanking enhancer sequence²⁴. Inclusion of two copies of this element results in increased response to serum stimulation²⁴. Therefore, the effect of EGF on gene expression was evaluated in cell lines expressing the K⁷²¹ or M⁷²¹ EGF receptor using a fusion gene containing two copies of the *c-fos* 5' regulated enhancer element (-357 to -276), the herpes simplex virus (HSV) thymidine kinase (TK) gene promoter (-200 to +70) and either the luciferase gene²⁵ or the bacterial chloramphenicol acetyl transferase (CAT) gene²⁶. EGF-induced stimulation was determined by the luciferase (or CAT) activity in cell extracts prepared after transient transfection of the fusion gene. B82 and CHO cells expressing the K⁷²¹ EGF receptor showed a 2.5- to 5.5-fold increase of reporter gene expression in response to EGF; however, cells expressing the M⁷²¹ receptor failed to show any EGF-dependence of reporter levels (Fig. 1*d*). It therefore



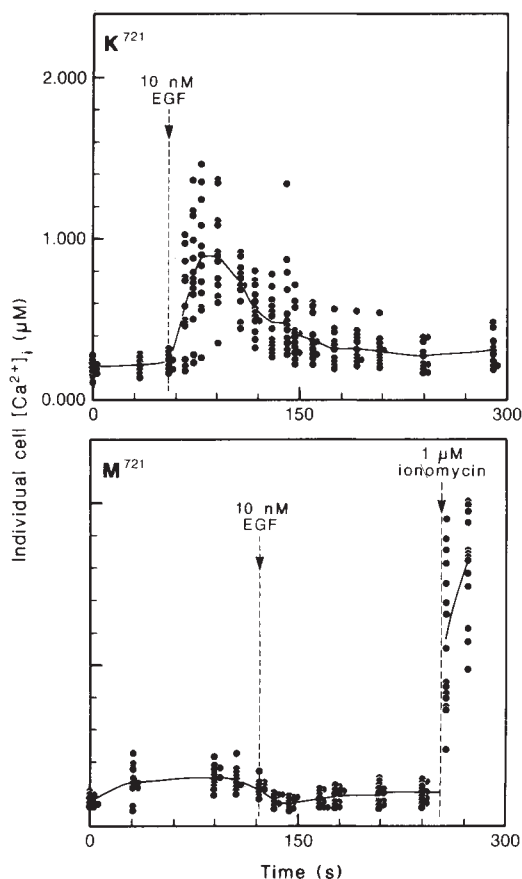


Fig. 2 Effects of EGF on rapid alterations of intracellular calcium. Fluorescence ratio imaging of individual cells was used to visualize the rapid effects of EGF on $[Ca^{2+}]_i$ in B82 cells expressing either K⁷²¹ (above) or M⁷²¹ (below) EGF receptors. Results of the fluorescent quantitation are converted to intracellular calcium, with the solid line representing the mean of individual cells for each time point. Similar results were obtained on repeated experiments. To confirm viability of the M⁷²¹ cells, 1 μ M ionomycin was added to demonstrate ability to increase $[Ca^{2+}]_i$.

Methods. Cells were grown in serum-containing medium on cover slips, then shifted to serum-free medium 48 h before experiments. Cells were loaded with fura-2 at 37 °C, and quantitation with the photometer was performed essentially as described^{28,29}.

seems that the rapid stimulation of *fos* gene expression by EGF is dependent on intact protein tyrosine kinase activity of the receptor.

One of the most rapid cellular responses induced by EGF is the increase, within seconds, of free intracellular calcium concentration ($[Ca^{2+}]_i$), considered a potential intracellular activator of the mitogenic pathway²⁷. Accurate continuous quantitation of rapid alterations in $[Ca^{2+}]_i$ in individual cells is possible using fluorescence imaging, utilizing fura-2 as a fluorescent indicator of $[Ca^{2+}]_i$ (refs 28 and 29). B82 cells expressing the K⁷²¹ EGF receptor showed a rapid response to addition of 10 nM EGF, with $[Ca^{2+}]_i$ increasing from a mean of 200 nM to 900 nM, reaching a plateau by 30 s, with a subsequent progressive decline. Treatment with 1 nM EGF resulted in similar kinetics of response, but with a submaximal increase of $[Ca^{2+}]_i$ (data not shown). B82 cells expressing the M⁷²¹ EGF receptor showed no EGF-dependent alterations in the basal $[Ca^{2+}]_i$, of 200 nM, even in the presence of 10 nM EGF, for up to 130 s (Fig. 2b). To confirm cell viability and dye responsiveness, a rapid, dramatic increase in $[Ca^{2+}]_i$ to $>1.5 \mu$ M was documented in response to the ionophore, ionomycin (Fig. 2b). Because influx of extracellular calcium and redistribution of internal stores may be mechanistically distinct events, the effect of EGF on intracellular calcium redistribution was evaluated in B82 cells express-

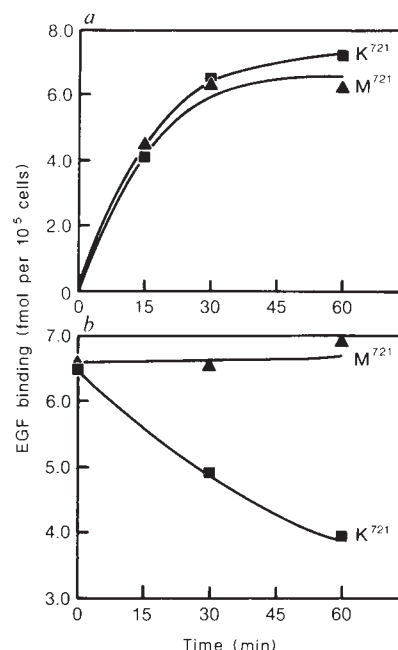


Fig. 3 Binding and down-regulation of K⁷²¹ and M⁷²¹ EGF receptors. **a**, B82 cells expressing K⁷²¹ (■) or M⁷²¹ (▲) receptors were analysed for kinetics of EGF binding as previously described²¹. Results are the average of triplicate determinations differing by $<5\%$. **b**, Kinetics of EGF receptor loss were quantitated following addition of 10 nM EGF for the indicated time periods. Bound EGF was removed by the washing procedures and binding of ¹²⁵I-labelled EGF measured, as previously described²¹. Receptor degradation in response to EGF has been confirmed by immunoprecipitation. Results are the average of triplicate determinations differing by $<5\%$.

ing K⁷²¹ EGF receptor in medium buffered to $<10^{-7}$ M extracellular Ca^{2+} . A small but statistically significant rapid increase in $[Ca^{2+}]_i$ (from 100 nM to 150 nM) was observed (data not shown). Therefore, EGF causes both Ca^{2+} influx and release from intracellular stores, and both responses must be lost to explain the non-responsiveness of $[Ca^{2+}]_i$ in the cells expressing the M⁷²¹ receptor.

The EGF receptor is rapidly internalized and degraded after binding its ligand; this internalization event begins within seconds or minutes after addition of EGF (ref. 30). In order to evaluate any potential relationship between activation of the intrinsic EGF receptor protein kinase activity and internalization events, the kinetics of binding and receptor down-regulation were evaluated in B82 and CHO cells expressing the M⁷²¹ and K⁷²¹ receptors. The kinetics of binding to both receptors was similar (Fig. 3a). As previously reported, the EGF receptors with either threonine or alanine at position 654 exhibit rapid internalization and progressive loss of EGF receptors in response to EGF (Fig. 3b and ref. 21); immunoprecipitation has confirmed that these receptors are degraded²¹. In contrast, homologous 'down regulation' of M⁷²¹ receptors is abolished even with prolonged exposure to high concentrations of EGF (Fig. 3b). The characteristic EGF-dependent loss of EGF receptor thus appears to require intact intrinsic protein tyrosine kinase activity, consistent with the decreased down-regulation observed with EGF receptors with carboxyl-terminal truncations³¹. This could reflect either the requirement for exogenous substrate phosphorylation or the allosteric effects of autophosphorylation of the carboxyl-terminus of the receptor³² for internalization and/or degradation.

A number of growth factor receptors with intrinsic protein tyrosine kinase activity seem to be essential for signalling a complex series of biochemical processes culminating in DNA

replication and cell division⁶. The effects of EGF on calcium mobilization, calcium flux, and gene transcription, as well as ultimate cell growth, appear to depend on intact protein tyrosine kinase activity. There is a formal possibility that ATP binding without phosphate transfer could be essential to some offshoot signalling mechanism. Recent data indicate that insulin-receptor-dependent glucose transport also requires intrinsic protein tyrosine kinase activity^{33,34}. We suggest that the intrinsic receptor protein tyrosine kinase activity is necessary for the described EGF-induced effects on cellular phenotype and receptor metabolism.

We thank Cliff Carpenter for aid and discussion; Charles Nelson for help with tissue culture; Harry Elsholtz for assistance; Eugene Gerner (University of Arizona) for the CHO cells used; and Suresh Subramani for the luciferase gene. These studies were supported by grants from NIH, the American Cancer Society, the Council for Tobacco Research, and the McKnight Foundation.

Received 16 June; accepted 24 June 1987.

1. Ushiro, H. & Cohen, S. *J. biol. Chem.* **255**, 8363–8365 (1980).
2. Downward, J. *et al. Nature* **307**, 521–527 (1984).
3. Ullrich, A. *et al. Nature* **309**, 418–425 (1984).
4. Lin, C. R. *et al. Science* **224**, 843–848 (1984).
5. Xu, Y.-H. *et al. Nature* **309**, 806–810 (1984).
6. Hunter, T. & Cooper, J. A. *Rev. Biochem.* **54**, 897–930 (1985).

7. Sawyer, S. T. & Cohen, S. *Biochemistry* **20**, 6280–6286 (1981).
8. Moolenaar, W. H., Yarden, Y., de Laat, S. W. & Schlessinger, J. *J. biol. Chem.* **257**, 8502–8506 (1982).
9. Murdock, G. H., Potter, E., Nicolaisen, A. K., Evans, R. M. & Rosenfeld, M. G. *Nature* **300**, 192–194 (1982).
10. Muller, R., Bravo, R., Burckhardt, J. & Curran, T. *Nature* **312**, 716–720 (1984).
11. Greensberg, M. C. & Ziff, E. G. *Nature* **311**, 433–438 (1984).
12. Kruijer, W., Cooper, J. A., Hunter, T. & Verma, I. *Nature* **312**, 711–716 (1984).
13. Kelly, K., Cochran, B. H., Stiles, C. D. & Leder, P. *Cell* **35**, 603–610 (1983).
14. Elsholtz, H. P. *et al. Science* **234**, 1552–1557 (1986).
15. Snyder, M. A., Bishop, J. M., McGrath, J. P. & Levinson, A. D. *Molec. cell. Biol.* **5**, 1772–1779 (1985).
16. Hanni, M. & Donoghue, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7894–7898 (1985).
17. Weinmaster, G., Zoller, M. J. & Pawson, T. *EMBO J.* **5**, 69–76 (1985).
18. Defize, L. H. K., Moolenaar, W. H., van der Saag, P. T. & de Laat, S. W. *EMBO J.* **5**, 1187–1192 (1986).
19. Prywes, R., Livneh, E., Ullrich, A. & Schlessinger, J. *EMBO J.* **5**, 2179–2190 (1986).
20. Sternberg, M. J. E. & Taylor, W. R. *FEBS Lett.* **175**, 387–392 (1984).
21. Lin, C. R. *et al. Cell* **44**, 839–848 (1986).
22. Simonsen, C. C. & Levinson, A. D. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2495–2499 (1983).
23. Gill, G. N. & Lazar, C. S. *Nature* **293**, 305–307 (1981).
24. Treisman, T. *Cell* **46**, 567–574 (1986).
25. deWet, J. R., Wood, K. V., DeLuca, M., Helinski, D. R. & Subramani, S. *Molec. cell. Biol.* **7**, 725–737 (1987).
26. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044–1051 (1982).
27. Moolenaar, W. H., Tertoolen, L. G. J. & de Laat, S. W. *J. biol. Chem.* **259**, 8066–8069 (1984).
28. Poenie, M., Alderton, J., Tsien, R. Y. & Steinhardt, R. A. *Nature* **315**, 147–149 (1985).
29. Poenie, M., Alderton, J., Steinhardt, R. & Tsien, R. *Science* **233**, 886–889 (1986).
30. Schlessinger, J., Shechter, Y., Willingham, M. C. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **75**, 2659–2663 (1978).
31. Livneh, E. *et al. J. biol. Chem.* **261**, 12490–12497 (1986).
32. Downward, J., Parker, P. & Waterfield, M. D. *Nature* **311**, 483–485 (1984).
33. Chou, C. K. *et al. J. biol. Chem.* **262**, 1842–1847 (1987).
34. Ebina, Y. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 704–708 (1987).
35. Weber, W., Gill, G. N. & Spiess, J. *Science* **224**, 294–297 (1984).
36. Bertics, P. J. & Gill, G. N. *J. biol. Chem.* **260**, 14642–14647 (1985).

Cell-cycle regulatory sequences in a hamster histone promoter and their interactions with cellular factors

Alexander Artishevsky*, Scott Wooden, Ajay Sharma, Elpidio Resendez Jr & Amy S. Lee†

Department of Biochemistry and the Norris Cancer Research Institute, University of Southern California School of Medicine, Los Angeles, California 90033, USA

Knowledge of how genes are regulated during the cell cycle is essential for understanding the process of cell growth on a molecular level. Numerous studies have established that, as mammalian cells go through the cell cycle, histone mRNA levels change, the largest amount being produced in the S phase^{1–7}. Both transcriptional and post-transcriptional mechanisms are responsible for this regulation and it has recently been demonstrated that nucleotide sequences in both the 5' and 3' termini of the histone gene are involved^{8–10}. From deletion analysis of a hamster *H3.2* fusion gene, we report here that the crucial control signals for both cell-cycle regulation and high level expression *in vivo* are contained in a 32-nucleotide (nt) region about 150 nt upstream of the TATA sequence and do not require any histone protein coding sequence. By comparison, the promoter of the herpes simplex virus (HSV) thymidine kinase gene is serum-stimulated but not cell-cycle regulated. The cell-cycle control exerted by the histone DNA regulatory element acts at the transcriptional level, as the rate of transcription is stimulated during the DNA synthetic phase of the cell cycle. Using DNA-protein mobility shift experiments, we demonstrate the existence of high affinity cellular factors interacting with the histone *H3.2* promoter sequence. The concentration of the protein-DNA complexes shows cell-cycle variation, particularly during the transition from late G₁ to the DNA synthesis phase. These data provide evidence for *in vivo* interactions between the cell-cycle transcriptional regulatory factors and the *cis*-acting DNA domain.

The 5' DNA sequence of the hamster *H3.2* gene (Fig. 1a) revealed the presence of a TATA and two CCAAT sequences. The consensus sequence GGGCGGG for the binding of the transcriptional factor Sp1 (ref. 11) occurs four times. Other features, such as G+C rich inverted repeats shown to be important for efficient transcription of histone genes *in vitro*^{12,13} and histone-specific sequences such as GGTCC¹⁴ and GACTTC¹² are also found in this highly G+C rich (68%) sequence extending to 426 nucleotides (nt) upstream of the translation initiation site. Interestingly, this hamster *H3.2* gene appears to be a homologue of the mouse *H3.2-614* gene¹⁵ (S.W. and A.S., unpublished results).

To examine the DNA sequence requirement for proper cell-cycle transcriptional regulation *in vivo*, we showed previously that a DNA fragment containing ~1 kilobase (kb) of a 5' flanking and 60 nt of the hamster histone *H3.2* coding sequence is able to confer cell-cycle regulation on a neomycin resistance gene (*neo*) *in vivo*^{8,16}. To define the regulatory domain, a series of 5' deletion mutants derived from the prototype histone-*neo* fusion gene pHN/7 was constructed by *Bal31* nuclease digestion (Fig. 1b). In this series of deletion mutants, decreasing lengths of the 5' flanking sequence including the coding sequence for the first 20 amino acids of *H3.2* were fused in the same transcriptional orientation to the *neo* coding region. Transfection of these plasmids by DNA-mediated gene transfer into the K12 cells^{17,18}, a temperature-sensitive (ts) mutant derived from Chinese hamster fibroblasts, Wg1A, led to the isolation of clonal lines containing the transfecting DNA. To eliminate bias towards transfectants which expressed *neo*, stable transfectants were selected on the basis of either G418 or HAT (hypoxanthine aminopterin and thymidine) resistance¹⁹. In the latter case, the deleted mutants were co-transfected into K12 cells with pSV2gpt (ref. 20). After expanding the cultures, we examined the pattern of *neo* expression during the cell cycle. Two approaches were used for this analysis. In one we synchronized the transfected K12 cells in G₁ phase by the serum deprivation method previously described^{18,21}. Addition of fresh medium led to the synchronous progression of these cells through the cycle, as shown by ³H-thymidine incorporation into the cellular DNA. The time course of these experiments was 23 hours, an interval equivalent to the doubling time of K12 cells in culture. At various times during the cell cycle, total cytoplasmic RNA was extracted

* Present address: Division of Biology, California Institute of Technology, Pasadena, California 91125, USA.

† To whom correspondence should be addressed.

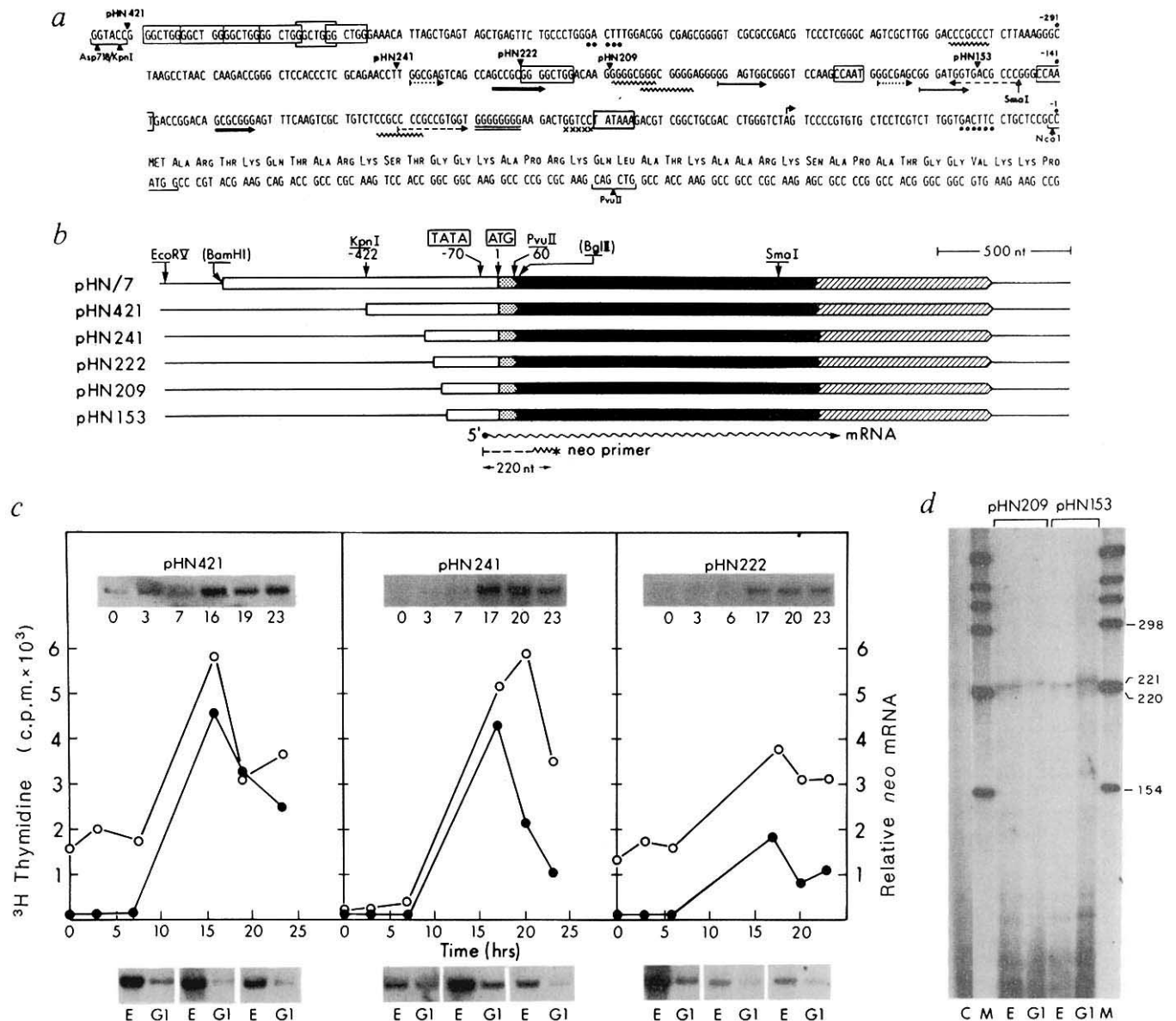


Fig. 1 Construction of the histone promoter deletion mutants and their regulation during the cell cycle. **a**, The 5' flanking and partial coding sequence of the hamster histone *H3.2* gene. The A residue in the ATG initiation codon is numbered as 1. Boxes, the TATA, CCAAT, and tandem repeat GGCTGG motifs. Right-angle arrow, the major transcriptional initiation site; double underline, poly G residues; bold, normal and dotted arrows, three sets of direct repeats; broken arrows, GC-rich inverted repeats; $\times \times \times$ and $\bullet \bullet \bullet$ underline, conserved histone-specific sequence (GGTCC)¹⁴ and (GACTTC)¹³ respectively; sawtooth underline, Sp1 consensus sequence¹¹. **b**, Structure of the parental plasmid pHN/7 and its derivative 5' deletion promoter mutants. Open bar, histone 5' flanking sequence; stippled bar, histone coding sequence; filled bar, neomycin resistance gene coding sequence; hatched bar, polyadenylation site; lines, pBR322 sequence. **c**, Cell-cycle analysis of the histone promoter deletion mutants pHN421, pHN241 and pHN222 by RNA blot hybridization³⁵. The relative levels of the *neo* transcripts determined by densitometry of the autoradiograms (O) shown in the insert were plotted with relative rates of DNA synthesis (●) at different times. Below, *neo* transcripts from individual K12 stable transfectants during exponential growth conditions (E), or in temperature-arrested cells (G₁). For pHN421 and pHN241 transfectants, three and six transfectants examined exhibited high and medium levels of *neo* transcripts, respectively. For pHN222 transfectants, three, eight and five transfectants examined exhibited high, medium and low levels of *neo* transcripts, respectively. **d**, Cell-cycle regulation of deletion mutants pHN209 and pHN153 assayed by primer extension. Lanes M, kinased size markers from pBR322 digested with *Hinf*I; lane C, 50 μ g of yeast transfer RNA was used as template for the primer extension reaction. Lanes E, exponentially growing cells; lanes G₁, cells arrested in mid G₁ phase by shifting to 40 °C.

Methods. The DNA was sequenced with Sanger's dideoxy chain termination method³³ using the M13 vector, with DNA fragments isolated from the plasmid pAAD3.7 (ref. 8). The 5' deletion end points were determined by the method of double-stranded sequencing³⁴ using as primer a 17-mer (5'TAGGCATAGGCTTGTT3') coding for the pBR322 sequence adjacent to the *Eco*RV site bordering the 5' end points. For construction of the deletion mutants, the parental plasmid pHN/7 was linearized with *Kpn*I; *Bal*31 was used to generate progressively shorter fragments of the *H3.2* promoter. To create an identical 5' border, the *Bal*31-treated plasmids were digested with *Eco*RV before self-ligation with T4 DNA ligase and subsequent transfection into HB101. The deleted plasmids were used to transfect K12 cells. Individual transfectants were isolated and the amounts of *neo* transcripts in synchronized cells were determined as previously described⁷. The rate of DNA synthesis was monitored by the incorporation of ³H-thymidine⁷. The *neo* transcript in pHN209 and pHN153 was determined by primer extension with a synthetic 23-mer oligonucleotide corresponding to the anti-sense strand of the *neo* gene²⁴. The extended products from equal amounts (50 μ g) of template RNA isolated from exponentially growing (E) or temperature-sensitive arrested (G₁) cells were analysed on a 6% 8 M urea DNA sequencing gel. The autoradiograms are shown.

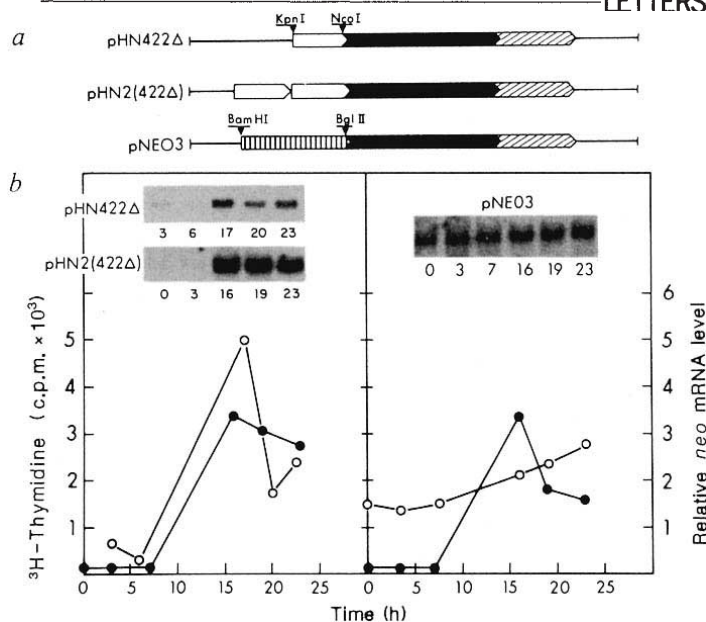


Fig. 2 *a*, Structure of pHN422Δ, pHN2(422Δ) and pNEO3. *b*, Cell-cycle regulation of the deletion plasmids which are void of histone coding sequence and the HSV TK promoter/*neo* fusion gene.

Methods. For the construction of pHN422Δ and pHN2(422Δ), a 422-nt *KpnI*-*NcoI* fragment isolated from pHN/7 and a 4.4-kb *BamHI*-*BglII* fragment isolated from plasmid pNEO3 (ref. 8) were treated with T4 DNA polymerase, ligated and transfected into *Escherichia coli* HB101. For the construction of pNEO3, a 800-nt *BamHI*-*BglII* promoter fragment of the HSV TK gene promoter was fused to the *neo* gene²⁵. The plasmids were transfected into K12 cells and stable transfectants were analysed for *neo* mRNA levels and rate of DNA synthesis in serum-synchronized cells as described in Fig. 1c legend. The quantification of the *neo* mRNA levels in pHN422Δ and pNEO3 transfectants is shown.

and hybridized to a *neo*-specific probe. Examples of the relative *neo* messenger RNA levels during the cell cycle for transfectants harbouring pHN421, pHN241 and pHN222 are shown in Fig. 1c. For the second approach we took advantage of the fact that the K12 cell line is a well characterized temperature-sensitive cell cycle mutant. By placing the transfected cells at 40 °C for 24 hours we obtained a population consisting mainly of cells arrested at the mid-G₁ stage of their cycle^{22,23}. As a control we grew cells exponentially at 35 °C; such cultures contained cells at various stages of the cell cycle. Total cytoplasmic RNA was extracted from these cells and hybridized to a radiolabelled *neo* probe (Fig. 1c, below).

Based on the belief that proper cell-cycle regulation of the *neo* gene would result in regulation resembling that of the histone, we expected: (1) low *neo* transcript levels in G₁ cells, followed by rapid increase in S phase cells and gradual decrease in late S and G₂ cells; and (2) lower *neo* transcript levels in temperature-arrested G₁ cells than exponentially growing cells. We observed that transfectants containing deletion mutants pHN421, pHN241 and pHN222 which contained 421, 241 and 222 nt respectively of the 5' flanking sequence exhibited cell-cycle regulation of the *neo* transcriptional unit. However, in the case of pHN222 transfectants, the general amount of *neo* mRNA appeared to have decreased and the differential *neo* RNA level between G₁ and S cells diminished in the majority of the transfectants. Therefore, in this construct, part of the transcriptional regulatory sequences required for maximal expression might have been deleted. Furthermore, transfection of K12 cells with deletion mutants pHN209 and pHN153 resulted in a 10-fold decrease in transfection efficiency for G418-resistant colonies, suggesting that the transcriptional activity of the remaining 209 nt of the 5' flanking sequence was much reduced. The *neo*

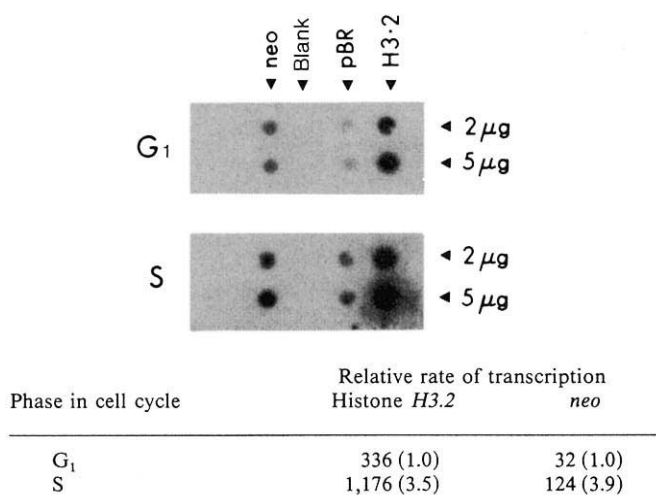


Fig. 3 *In vitro* transcription rates of endogenous H3.2 and the transfected *neo* genes. The autoradiograms were quantified by densitometry and the results are summarized below. Numbers in parentheses indicate increase in the rate of transcription in S phase as compared to G₁.

Methods. DNA for each plasmid: (7 μg pBR322, pNEO3, pAAD3.7 encoding the plasmid vector, *neo* and hamster H3.2 respectively) was denatured by incubation in 0.5 M NaOH for 20 min at 22 °C. After neutralizing with HCl, the DNA (2 or 5 μg) was adjusted to the total volume of 100 μl with 15×SSC (1×SSC: 0.15 M NaCl, 0.015 M sodium citrate) and the DNA was applied to nitrocellulose paper (presoaked in 5× and 20×SSC and air dried). Blank dots were prepared by adding 100 μl of 15×SSC to the appropriate wells. Wells were rinsed with 100 μl of 15×SSC. The filters were air dried and baked at 80 °C for 2 h. Before hybridization, the filters were incubated in 10×Denhardt's (1×Denhardt's is 0.02% each bovine serum albumin (BSA), polyvinyl pyrrolidone (PVP) and Ficoll), 0.1% SDS and 250 μg heat denatured salmon sperm DNA for 1–2 h at 42 °C. The radiolabelled nuclear RNA was prepared from synchronized pHN/7–15 transfectants at G₁ and S phase (2 h and 15 h after addition of fresh medium, respectively) as previously described^{7,8}. *In vitro* labelled nuclear RNA (6 and 2.3 × 10⁷ c.p.m. isolated from equal numbers of G₁- and S-phase nuclei) was added to 10 ml hybridization buffer (50% deionized formamide, 5×SSC, 1×Denhardt's, 0.02 M phosphate buffer (PB), 0.1% SDS, 250 μg of heat denatured salmon sperm DNA) and incubation was continued for 16 h at 42 °C. The blots were washed once at 22 °C for 40 min in 4×SSC in washing buffer (0.1% sodium pyrophosphate, 0.025% PB, 0.1% SDS), twice in 1×SSC in the same buffer at 50 °C for 45 min, and once in 0.3×SSC in the same buffer at 22 °C for 30 min: autoradiograms are shown.

mRNA in these transfectants was also undetectable in RNA blot hybridizations (data not shown). When we analysed these transfectants by the more sensitive primer extension method using a *neo*-specific oligonucleotide²⁴, we found that small amounts of properly initiated *neo* transcripts could be detected; however, the level of *neo* in exponentially growing cells was about the same as that of G₁ cells (Fig. 1d). Therefore, whereas deletions pHN209 and pHN153 maintained a low level of *neo* expression, they had lost the cell-cycle regulation observed in pHN421, pHN241 and pHN222. Both G418- and HAT-resistant transfectants showed similar patterns of *neo* expression. Based on these results, we conclude that the 32-nt region between deletions pHN241 and pHN209 is crucial for both cell-cycle regulation and high-level expression, and that the method of transfectant selection and site of integration are not important as long as enough 5' sequence is included in the fusion gene.

In all the deletion mutants analysed above, 60 nt of the histone coding sequence were still present. To eliminate this sequence, the two plasmids pHN422Δ and pHN2(422Δ), were constructed (Fig. 2a). They contain respectively one and two copies of the

histone promoter fused to the *neo* transcriptional unit. Following transfection of these plasmids into K12 cells and selection for either G418 or HAT resistance, the expression of *neo* mRNA during the cell cycle was examined in stable transfectants (Fig. 2b). Stringent regulation was observed in both serum-synchronized cell cultures (Fig. 2b) and in G₁ cells arrested by the K12 temperature-sensitive mutation (data not shown). Interestingly, the level of *neo* mRNA was generally higher in pHN2(422Δ) transfectants containing two copies of the histone promoter. Thus, we conclude that internal histone coding sequence is not necessary for cell-cycle regulation.

To discover whether other promoters are capable of conferring cell-cycle regulation to the *neo* gene fused downstream, we obtained G418-resistant stable transfectants of the parental plasmid pNEO3 (ref. 25), which contains the HSV *TK* gene promoter. The expression of *neo* mRNA during the cell cycle under the control of the HSV *TK* promoter was different from that observed for the *H3.2* promoter. Instead of the sharp increase during S and decline in G₂ observed for the *H3.2* promoter, there was a slight but gradual increase in *neo* transcript as cells entered S phase, even after DNA replication had subsided (Fig. 2b). This indicated that the HSV *TK* promoter was sensitive to serum growth stimulation rather than cell-cycle regulation.

Next we analysed the transcriptional rates of the *neo* gene in these stable transfectants during G₁ and S phases. Transfectants harbouring the pHN/7 plasmid⁸ were grown up and nuclei were isolated from equal numbers of G₁- and S-phase cells and labelled *in vitro* with [α -³²P]UTP. The relative rates of transcription were measured by hybridization of the labelled nuclear RNA with excess *neo* plasmid DNA bound on nitrocellulose filters⁷. The autoradiograms (Fig. 3) were quantitated by densitometry. Our results showed that the relative increase in the transcriptional rates during S phase for the endogenous *H3.2* gene and the histone-*neo* fusion gene were similar. The 3- to 4-fold increase in the transcriptional rate of the endogenous *H3.2* gene was consistent with previous measurements in hamster and mouse fibroblasts using both *in vivo* and *in vitro* labelling methods⁵⁻⁷. These measurements were also performed with pHN241 transfectants synchronized by the K12 temperature-sensitive mutation and similar results were obtained (data not shown). Based on these results, we conclude that the cell-cycle regulation of the histone-*neo* fusion gene is at least partly attributable to regulation at the transcriptional level. A comparative analysis of the DNA sequence within and bordering the 32-nt region (Fig. 1a) essential for the efficient use of the *H3.2* promoter in cell-cycle transcriptional regulation revealed several G+C rich motifs and sequences resembling GCGAAA, a sequence motif found in the yeast histone and *HO* genes which are also cell-cycle regulated^{26,27}. Interestingly, the sequence GGCTGG which is repeated six times in the upstream region is also in this 32-nt region. Except for a few base changes, this region is highly conserved compared with the homologous mouse *H3.2-614* sequence.

The simplest mechanism which can explain the transcriptional regulation observed is that the cell-cycle regulating sequence interacts with some positive factors to activate histone mRNA transcription. The stringent cell-cycle regulation observed in stable transfectants reported here and elsewhere^{8,28,29} indicates that diffusible factors are available to regulate the transfected histone genes. As a first step, we searched for cellular factors that would bind with high affinity to the *H3.2* promoter using 'protein distribution analysis'³⁰. Whole cell protein extracts were prepared from synchronized cells at various times after the addition of fresh medium to the serum-starved K12 cells and mixed with labelled *KpnI-NcoI* fragment of the *H3.2* promoter. To enhance the sensitivity of this assay, a simple alternating copolymer duplex poly(dI-dC)·poly(dI-dC) was used to bind the more abundant nonspecific DNA-binding proteins in the extract³¹. The mixtures were electrophoresed on low salt gels and the labelled DNA was detected by autoradiography

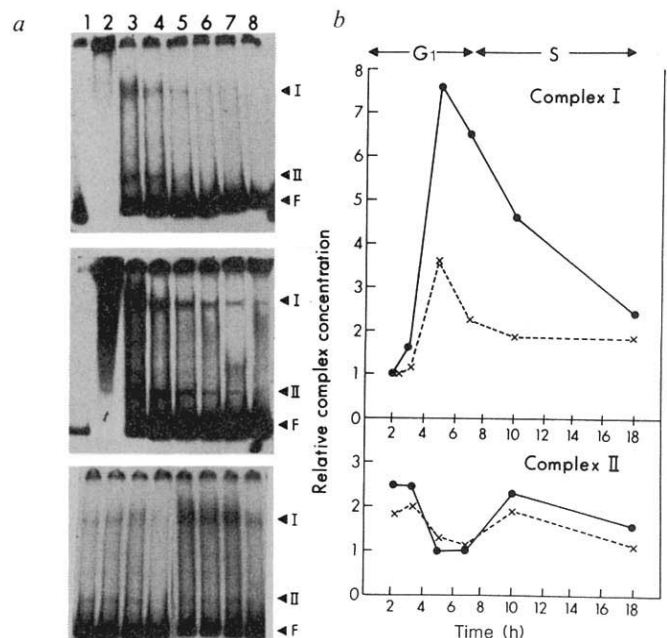


Fig. 4 Binding of cellular factors to the hamster histone *H3.2* promoter. *a*, Gel electrophoresis DNA binding assay with the labelled histone *H3.2* promoter (Asp 718/*NcoI* fragment) and whole cell protein extracts prepared from serum-synchronized K12 cells 2 h (top panel 1) and 6 h (middle and lower panels) after addition of fresh serum. In the top two panels, the titration profile with increasing amounts of poly(dI-dC)·poly(dI-dC) is shown. The protein extract was omitted in lane 1. Lanes 2-8 contained 0, 40, 100, 200, 400, 800 and 1,600 ng of poly(dI-dC)·poly(dI-dC) respectively. In the bottom panel, the competition of the complex with unlabelled promoter fragment and poly(dI-dC)·poly(dI-dC) is shown. In addition to the 200 ng of poly(dI-dC)·poly(dI-dC) present in the binding reactions, lanes 1-4 contained 2, 4, 10 and 20 ng of the unlabelled Asp 718/*NcoI* fragment respectively; lanes 5-8 contained 10, 20, 40 and 400 ng of poly(dI-dC)·poly(dI-dC). The positions of free (F) and complexed fragments (I and II) are indicated. *b*, Relative concentrations of complex I and II during the cell cycle. Equal amounts of protein extract (1 µg) from synchronized K12 cells progressing 2, 3, 5, 7, 10 and 18 h into the cell cycle were incubated with the labelled Asp 718/*NcoI* fragment in presence of 50 ng (●) and 100 ng (×) of poly(dI-dC)·poly(dI-dC). The relative complex concentrations quantitated by densitometry of the autoradiograms were plotted against the time during the cell cycle when the protein extracts were prepared.

Methods. The Asp 718/*NcoI* fragment was labelled with [α -³²P]-dATP and T4 DNA polymerase. The protein extracts were prepared by the method of Manley *et al.*³⁶. The binding reaction (20 µl) contained 1 ng (~8,000 c.p.m.) of the labelled fragment, 10 mM Tris-HCl (pH 7.5), 50 mM NaCl, 1 mM EDTA, 10 mM β-mercaptoethanol and 0.5 µg (top panel) and 1 µg (lower two) whole cell protein extract. For the competition experiment, 1 ng (~2,000 c.p.m.) of the labelled fragment was used. After incubation at room temperature for 20 min, the resulting complexes are resolved in a low-ionic strength 5% polyacrylamide gel (acrylamide/bisacrylamide, weight ratio of 80:1) containing 6.7 mM Tris-HCl (pH 7.5), 3.3 mM sodium acetate and 1 mM Na-EDTA^{30,37}. With buffer recirculation, the gel was pre-electrophoresed for 2 h at 9 V cm⁻¹ and the samples were electrophoresed at the same voltage gradient for 2 h. The gel was dried and autoradiographed at -70°C with an intensifying screen. The autoradiograms were quantitated by densitometry.

(Fig. 4a). In the absence of competitor DNA, all the labelled fragments were retained at the top of the gel, probably because of binding of an excess of non-sequence-specific proteins. With addition of increasing amounts of poly(dI-dC)·(dI-dC) as competitors, two major protein-DNA complexes (I and II) which migrated more slowly than the free fragment were detected. The relative amounts of these complexes were quantitated by densitometry and plotted against the time during the cell cycle when the extracts were prepared (Fig. 4b). Our results indicated that at least two cellular factors were able to bind to the *H3.2* promoter with high affinity. Based on a competition assay using unlabelled *H3.2* promoter fragment and assuming a single affinity site within this 422-nt fragment, the factor or factors responsible for complex I and II have a 3.6×10^4 and 1.4×10^4 times higher affinity respectively for the cognate sequence than for the nonspecific polymer³¹ (Fig. 4a). Because for an equal amount of protein extract and in the presence of equal amounts of competitor DNA, the largest amount of radio-labelled fragment in complex I was observed for late G₁-phase (5 h) extracts, in this limited analysis the number of factor(s) involved in complex I was highest during the transition from G₁ to S phase, as DNA synthesis and the transcription of histone genes were about to start⁷. In contrast, the factor(s) involved in complex II appeared to bind to the histone promoter throughout the cell cycle, with a slightly lower abundance in late G₁ cells. It is possible that some of these factors are part of the general transcriptional machinery analogous to the common factor which interacts with the histone H2B as well as the immunoglobulin and small nuclear RNA promoters³², or they are cell-cycle specific regulatory molecules. Further characterization of the interactions of the purified cellular factors with the *H3.2* promoter may enable us to dissociate the components required for cell-cycle regulation from basal level transcription.

We thank Robert Maxson for review of the manuscript and discussions on protein-DNA interactions, and Steven Wells for technical assistance. E.R. is supported in part by a predoctoral

fellowship from the NIH. This work was supported by a PHS Grant from the NIH. A.S.L. is a recipient of an American Cancer Society Faculty Research Award.

Received 5 May; accepted 11 June 1987.

- Schumperli, D. *Cell* **45**, 471-472 (1986).
- Plumb, M., Stein, J. & Stein, G. *Nucleic Acids Res.* **11**, 2391-2410 (1983).
- Sittman, D. B., Graves, R. A. & Marzluff, W. F. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1849-1853 (1983).
- Heintz, N., Sive, H. L. & Roeder, R. G. *Molec. cell. Biol.* **3**, 539-550 (1983).
- DeLisle, A. J., Graves, R. A., Marzluff, W. F. & Johnson, L. F. *Molec. cell. Biol.* **3**, 1920-1929 (1983).
- Alterman, R.-B. M. *et al. Molec. cell. Biol.* **4**, 123-132 (1984).
- Artishevsky, A., Delegeane, A. M. & Lee, A. S. *Molec. cell. Biol.* **4**, 2364-2369 (1984).
- Artishevsky, A., Graftsky, A. & Lee, A. S. *Science* **230**, 1061-1063 (1985).
- Lüscher, B., Stauber, C., Schindler, R. & Schumperli, D. *Proc. natn. Acad. Sci. U.S.A.* **82**, 4389-4399 (1985).
- Stauber, C., Lüscher, B., Eckner, R., Lotscher, E. & Schumperli, D. *EMBO J.* **5**, 3297-3303 (1986).
- Kadonaga, J. T., Jones, K. A. & Tjian, R. *Trends biochem. Sci.* **11**, 20-23 (1986).
- Hanley, S. M., Bleecker, G. C. & Heintz, N. *Molec. cell. Biol.* **5**, 380-389 (1985).
- Sive, H. L., Heintz, N. & Roeder, R. G. *Molec. cell. Biol.* **6**, 3329-3340 (1986).
- Clerc, R. G., Bucher, P., Strub, K. & Birnstiel, M. L. *Nucleic Acids Res.* **11**, 8641-8657 (1983).
- Taylor, D., Wellman, S. E. & Marzluff, W. F. *Molec. Evol.* **23**, 242-249 (1986).
- Davis, J. & Smith, D. I. *A. Rev. Microbiol.* **32**, 469-518 (1978).
- Melero, J. A. *J. cell. Physiol.* **98**, 17-30 (1979).
- Delegeane, A. M. & Lee, A. S. *Science* **215**, 79-81 (1982).
- Wang, M.-L. J. & Lee, A. S. *Biochem. biophys. Res. Commun.* **110**, 593-601 (1983).
- Mulligan, R. C. & Berg, P. *Proc. natn. Acad. Sci. U.S.A.* **78**, 2072-2076 (1981).
- Dulbecco, R. *Nature* **227**, 802-806 (1970).
- Melero, J. A. & Fincham, V. J. *cell. Physiol.* **95**, 295-306 (1978).
- Ashihara, J., Cheng, S. D. & Baserga, R. J. *cell. Physiol.* **96**, 365-369 (1978).
- Lin, A. Y., Chang, S. C. & Lee, A. S. *Molec. cell. Biol.* **6**, 1235-1243 (1986).
- Attenello, J. W. & Lee, A. S. *Science* **226**, 187-190 (1984).
- Osley, M. A., Gould, J., Kim, S., Kane, M. & Hereford, L. *Cell* **45**, 537-544 (1986).
- Nasmyth, K. *Cell* **42**, 225-235 (1985).
- Capusso, O. & Heintz, N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 5622-5626 (1985).
- Alterman, R.-B. M., Sprecher, C., Graves, R., Marzluff, W. F. & Skoultschi, A. I. *Molec. cell. Biol.* **5**, 2316-2324 (1985).
- Fried, M. G. & Crothers, D. M. *J. molec. Biol.* **172**, 241-262, 263-282 (1984).
- Singh, H., Sen, R., Baltimore, D. & Sharp, P. A. *Nature* **319**, 154-158 (1986).
- Sive, H. L. & Roeder, R. G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6382-6386 (1986).
- Sanger, F., Nicklen, S. & Coulson, A. K. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).
- Guidelines for *Quick and Simple Plasmid Sequencing* (Boehringer Mannheim, 1986).
- Lee, A. S., Delegeane, A. M., Baker, V. & Chen, P. C. *J. biol. Chem.* **258**, 597-603 (1983).
- Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gefter, M. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 3855-3859 (1980).
- Strauss, F. & Varshavsky, A. *Cell* **37**, 889-901 (1984).

Erythroid-specific transcription of the chicken histone H5 gene is directed by a 3' enhancer

Cecelia D. Trainor, Susan J. Stamler
& James Douglas Engel*

Department of Biochemistry, Molecular Biology and Cell Biology,
Northwestern University, Evanston, Illinois 60201, USA

The expression of a variety of vertebrate genes is transcriptionally regulated through tissue-specific cellular enhancer elements¹. An erythroid-specific enhancer sequence has recently been identified 3' to (downstream of) the chicken adult β -globin gene^{2,3}. Here we report the identification of a second erythroid-specific enhancer sequence, which is 3' to the chicken histone H5 gene. The similarity of these two enhancer elements, with respect both to function and location relative to the target gene, implies some functional conservation in their evolution and in the mechanism by which they affect erythroid gene transcription.

Chicken histone H5 is an interesting candidate gene for analysis of early erythropoiesis for a number of reasons. First, histone H5 is a tissue-specific protein found only in species of animals harbouring nucleated erythrocytes (birds, fish, amphibians and reptiles)⁴, and is a replacement variant linker histone which gradually substitutes for histone H1 as erythroid cells mature⁵.

Second, histone H5 is among the earliest identified differentiation markers for the erythroid lineage. It is expressed in the earliest erythroid cells which can be tested, including embryonic and adult erythroid precursor cells arrested at the CFU-E stage by transformation with avian erythroblastosis virus (AEV) (refs 6-8 and unpublished data). Both H5 protein^{9,10} and its messenger RNA (unpublished data) are present in normal reticulocytes in larger amounts than in AEV-transformed progenitor cells.

The histone H5 gene of the chicken has been isolated as both complementary and genomic DNA¹¹⁻¹⁴. Histone H5 mRNA is not detectable by RNA blot analysis in poly(A)⁺ RNA isolated from liver, embryonic brain or fibroblasts (data not shown). To determine the molecular mechanism that regulates this tissue-specific mRNA accumulation, a marked H5 gene, pH5*, was introduced into erythroid and non-erythroid cells by transfection using DEAE-dextran or calcium phosphate^{2,15}. HD6 (an erythroid precursor cell line transformed with ts167 AEV; refs 15 and 16) was chosen as the initial transfection recipient cell because H5 mRNA levels are slightly higher in this line than in other tested erythroid precursor lines. HD3 is an erythroid precursor cell line transformed with ts34 AEV (ref. 17) and HD3 cells were used in these studies because they can be induced to mature terminally^{2,7}. Primary chicken embryo fibroblasts (CEF) were chosen as the non-erythroid control cells for these transfection studies.

The H5 gene was marked by intragenic insertion of an 8-base pair (bp) *Clal* linker to form pH5*. Thus, mRNA transcribed from this plasmid can be distinguished from the endogenous H5 transcript in erythroid cells. A diagram of the marked 2.5-kilobase (kb) H5* gene plasmid and its deletion mutants is

* To whom correspondence should be addressed.

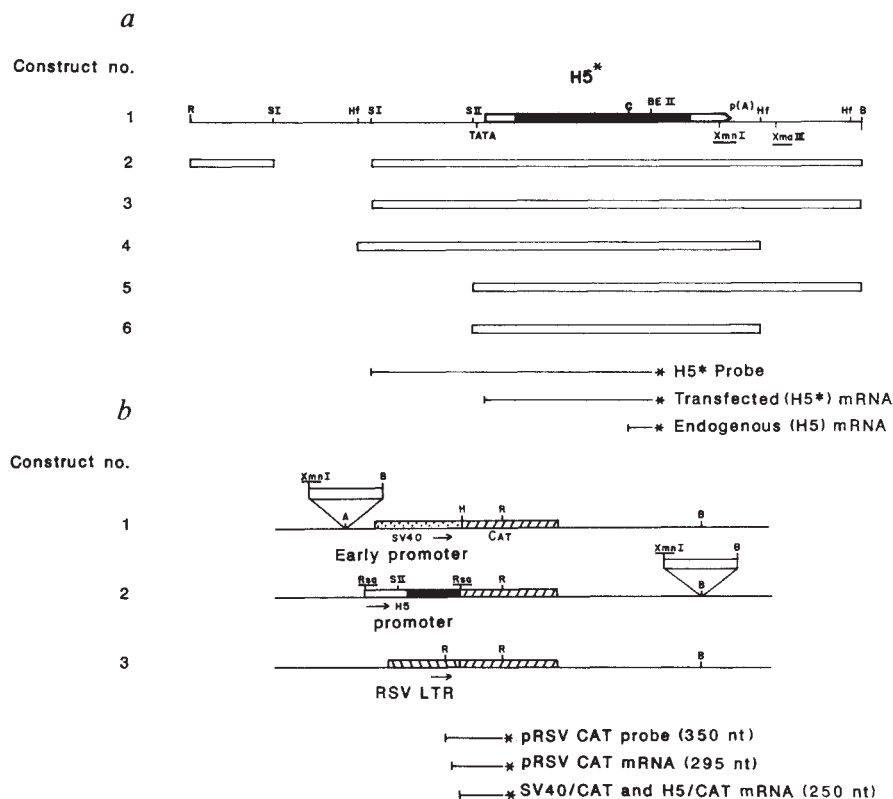


Fig. 1 *a*, Construction of the marked H5 gene, H5*, and deletion mutants. Top line, restriction map of a 2.5-kb genomic clone of chicken erythrocyte histone H5 modified by the insertion of an 8-bp *Cla*I linker at a unique, polymorphic¹⁶ *Sma*I site in the H5 locus to generate H5*. Boxes below depict a series of deletion mutants of H5* in which are removed all 5' flanking sequences to position -47 (ref. 15) and all 3' flanking sequences to within 114 bp of the polyadenylation site. Bottom, the H5* *S*₁ nuclease protection probe (a *Bst*EII-*Sac*I fragment) and expected protection fragments. Abbreviations: R, *Eco*RI; SI, *Sac*I; Hf, *Hin*FI; SII, *Sac*II; C, *Cla*I (linker); BEII, *Bst*EII; B, *Bam*HI; p(A), histone H5 mRNA polyadenylation site; TATA, conserved 5' element. *b*, Constructs containing the H5 gene 3' *Xmn*I-*Bam*HI fragment. Top line, pSV2Cat (ref. 20) was modified by the insertion of the H5 *Xmn*I-*Bam*HI fragment at the *Acc*I site 5' to the SV40 early promoter. Second line, pSVOCat (ref. 20) was modified by the insertion of a 600-bp H5 *Rsa*I fragment (Fig. 1*a*, top line) containing the H5 promoter, into the *Hind*III site 5' to the CAT gene (pSVO-H5 *Rsa*). Third line, this construction was further modified by the insertion of the 3' H5 *Xmn*I-*Bam*HI fragment at a *Bam*HI site 5' to the SV40 polyadenylation signal. The panel below the constructions shows the CAT *S*₁ nuclease probe and expected protection products.

shown in Fig. 1*a*. The *S*₁ nuclease probe generates a 72-base protection fragment from the endogenous H5 message and a 506-base protection fragment from a plasmid-derived transcript. As a control for transfection efficiency, an equal molar ratio of pRSVCat (ref. 18), which contains the chloramphenicol acetyl transferase (CAT) gene, was cotransfected in each experiment.

An *S*₁ nuclease protection assay of HD6 and CEF RNAs purified from cells transfected with pH5* and pRSVCat is shown in Fig. 2*a*. As expected, a 300-base CAT mRNA protection fragment is present in both transfected HD6 and CEF cells (Fig. 2*a*, lanes 2, 3). A 72-base (endogenous) H5 mRNA protection fragment is seen in RNA isolated from HD6 cells with or without transfection (Fig. 2*a*, lanes 1 and 2), but the 506-base H5* fragment is present only in transfected HD6 cells (Fig. 2*a*, lane 2). Neither endogenous (H5) nor transfected (H5*) mRNA is detected in transfected CEF (Fig. 2*a*, lane 3). The histone H5 gene is therefore expressed in a tissue-specific manner when introduced as an episome, and thus the erythroid-specific regulation of this gene must involve *cis*-acting DNA sequence(s) in pH5*.

To identify regulatory sequences in the H5 locus, a series of deletion mutants of pH5* were examined in transfection assay (Fig. 1*a*). All 5' flanking sequences to position -47 (relative to the H5 mRNA CAP site¹³), as well as 3' sequences to within 114 bp of the polyadenylation site, were removed in sequential fashion. This analysis indicates that sequences 5' to position -396 are not required for H5 mRNA accumulation in erythroid cells, because in mutants lacking these sequences the gene is expressed at the same level as the pH5* gene (Fig. 2*b*, lanes 1, 2, 3). In contrast, removal of either the remaining 5' or 3' regions of the locus results in a reduced level of expression (Fig. 2*b*, lanes 4, 5*a*) relative to pH5* after transfection into HD6 cells.

To define further the role of these putative regulatory regions, deletion mutant constructs 5 and 6 were compared in the transfection assay. Whereas mutant 5 is consistently expressed at about twofold lower levels than pH5* (Fig. 2*b*, lane 5*b*), mutant 6, which is identical to 5 except that it lacks the putative 3' regulatory region, is inert as a transcription substrate in HD6

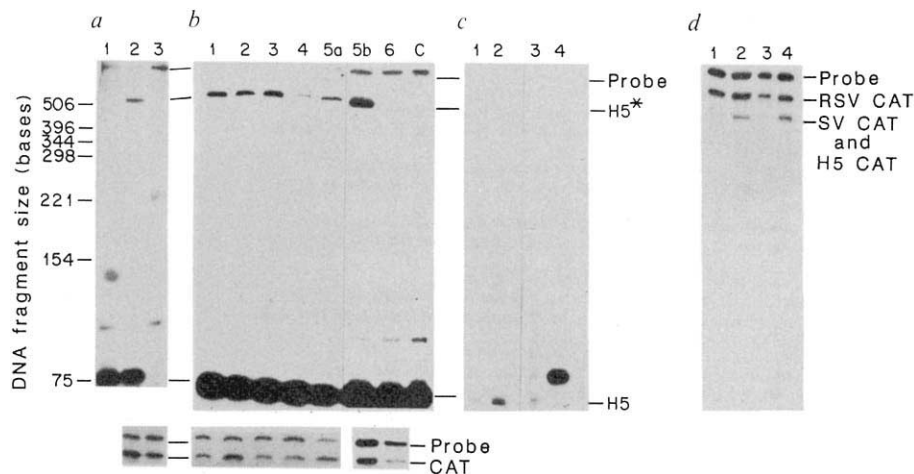
cells (Fig. 2*b*, lane 6). We conclude that the 5' and 3' deletions are each detrimental to H5* mRNA accumulation in HD6 cells, and that the combined effect of introducing both 5' and 3' deletions into the same construct is more deleterious than a simple additive or multiplicative accumulation of the two effects.

Histone H5* deletion mutants 2 to 6 were also cotransfected into CEF, with pRSVCat again serving as the positive transfection control. The CAT mRNA was easily detectable in each transfection reaction, whereas H5* mRNA was absent in every case (data not shown). Therefore, the 5' and 3' flanking sequences which have been removed from the H5 locus in the deletion mutants shown in Fig. 1*a* do not harbour a simple, singular silencer element which prevents H5 transcription in non-erythroid cells.

Histone H5 mRNA is more abundant in circulating erythroid cells of the chicken than in AEV-transformed erythroid precursor cells. Because we wished to determine whether the putative regulatory regions of the H5 locus are involved in the temporal as well as the tissue-specific regulation of H5 expression, the amount of endogenous H5 mRNA in differentiation uninduced (35 °C) and induced (42 °C) HD3⁷ cells was examined. The steady-state amount of H5 mRNA increased fourfold after differentiation (Fig. 2*c*, lanes 1 and 2), whereas the adult β -globin transcript, as expected², was much more strongly induced by the temperature shift (Fig. 2*c*, lanes 3 and 4). Therefore, induction of differentiation in HD3 cells is a useful model for analysis of histone H5 temporal regulation and tissue-specificity.

The 5' and 3' regulatory regions identified by the deletion experiments were further analysed by insertion into the T-antigen 'enhancer trap'¹⁹. Enhancer activity is detected by immunofluorescence analysis using monoclonal antibody against T antigen². It became clear that T-antigen accumulation is strongly dependent upon the 527-bp 3' *Xmn*I-*Bam*HI histone H5 fragment in either orientation in HD3 cells, whereas no differential T-antigen accumulation was detectable, comparing the same constructs, in CEF (data not shown). Thus the H5 *Xmn*I-*Bam*HI fragment harbours a tissue-specific enhancer element. Furthermore, both the intensity of T-antigen fluores-

Fig. 2 a, The histone H5 gene is tissue-specifically transcribed in transient transfection assay. HD6 cells and CEF were cotransfected with equal molar ratios of pH5* and pRSVCat as described previously^{2,15}. Total cellular RNAs were purified by lysing cells in guanidinium isothiocyanate and pelleting through CsCl (refs 21 and 22). Nuclease S₁ protection assays were performed according to published procedures²³. Lane 1, 20 µg of HD6 RNA, lane 2, 20 µg of HD6 RNA from cells cotransfected with pH5* and pRSVCat, lane 3 is 30 µg of CEF RNA from cells cotransfected with the same two plasmids. Upper panel, results with the H5* probe; lower panel, CAT probe controls for transfection efficiency. **b**, H5* deletion mutants localize two potential *cis*-acting regulatory DNA regions in the H5 locus. Nuclease S₁ protection assays of HD6 RNA from cells cotransfected with pRSVCat and one of the H5* deletion mutant constructions (see Fig. 1a) were performed. Lane numbers correspond to construct numbers from Fig. 1a. Lanes 1 to 5a are from one experiment; lanes 5b and 6 are from a separate experiment. Lane C shows the protection of untransfected HD6 cell RNA. **c**, The steady-state level of histone H5 mRNA is about four times higher in differentiation-induced HD3 cells. Nuclease S₁ analysis of RNAs from HD3 cells propagated at 35 °C or 42 °C in differentiation media² are shown. Lane 1, 2 µg of uninduced (35 °C) HD3 cell RNA; lane 2, 2 µg of RNA from HD3 cells induced to differentiate for 42 h, both tested with the H5* probe. Lanes 3 and 4, equal amounts of the adult β-globin transcript in the 35 °C and the 42 °C RNA populations, respectively, as a control for differentiation induction. The β*-globin probe and S₁ nuclease conditions used are as previously described². **d**, The histone H5 3' *Xmn*I–*Bam*HI fragment harbours an enhancer element. Plasmids pSV2Cat and pSVO/Rsa H5 (with and without the 3' H5 *Xmn*I–*Bam*HI fragment; see Fig. 1b, constructs 1 and 2) were cotransfected into HD3 cells with pRSVCat. RNA (30 µg) from differentiation-induced cells was assayed by S₁ nuclease protection using the 350-bp *Eco*RI probe from pRSVCat (Fig. 1b, construct 3). Lanes 1 and 2, results with pSV2Cat without and with the H5 527-bp fragment, respectively; lanes 3 and 4 show analogous results with pSVO/H5Rsa without and with the H5 enhancer element, respectively. Similar results to those shown in lanes 1 and 2 were seen when the SV40 enhancer was removed from the construction (Fig. 1b, construct 1), implying that the SV40 enhancer is inactive in chicken erythroid cells (data not shown).



cence and the proportion of positive cells increased significantly in the transfected HD3 cells propagated under differentiation-inducing conditions relative to the uninduced cell population, indicating that the H5 enhancer also mediates the temporal induction of H5 mRNA during erythroid cell maturation (data not shown). The 527-bp 3' histone H5 enhancer fragment was further subdivided into 215- and 312-bp fragments by digestion with *Xma*III (Fig. 1a). The 215-bp *Xmn*I–*Xma*III fragment proximal to the gene retains enhancer activity in the T-antigen assay, whereas the larger (*Xma*III–*Bam*HI) fragment is not active. The H5 5' flanking region (*Sac*I–*Sac*II, Fig. 1a) was inactive in this assay.

To confirm that the H5 enhancer exerts a direct transcriptional effect on the Simian virus 40 (SV40) early promoter and on its own promoter, two sets of constructs were prepared and transfected into HD3 cells: one contained the bacterial CAT gene transcriptionally directed by the SV40 early promoter (with and without the H5 gene 3' 527-bp *Xmn*I–*Bam*HI fragment), whereas the other contained the same gene directed by the H5 promoter with and without the H5 enhancer (Fig. 1b). CAT gene transcripts are easily detectable with both constructs when the 3' H5 gene 527-bp enhancer fragment is present, whereas these RNAs are ~10-fold less prevalent when transcribed by only the SV40 promoter or only the H5 promoter (Fig. 2d). Furthermore, the 3' H5 enhancer functions equally well when 3' or 5' to the test gene (Fig. 1b). Thus the 3' histone H5 gene 527-bp *Xmn*I–*Bam*HI fragment fulfils the criteria for an enhancer element: it is able to potentiate expression of a foreign gene from a foreign promoter in a position- and orientation-independent fashion. Curiously perhaps, the enhancer does not show any preference for its own promoter over that of a monkey viral promoter.

Comparison of the histone H5 enhancer sequence in the 215-bp *Xmn*I–*Xma*III fragment (Fig. 1a) to the β-globin 3' enhancer^{2,3} reveals a striking nucleotide sequence similarity between the two: a 34-bp segment was found in both enhancers which has 8 nucleotide mismatches (shown in lower case in the β-globin enhancer sequence) and a single base pair nucleotide gap (dash):

H5E: 5' GGAGGAGAGGGGACTCCTTCTGTCCATAGGAGT
βE: 5' GGAaGAGAGGGGgtTaaTcC–TGTCaATAGtAGT

A region of similarity such as this has a probability of occurring by chance in two unrelated segments of DNA of less than one in 10¹¹, and therefore we anticipated that this DNA segment might be important physiologically in erythroid cell-specific gene regulation. To determine whether this sequence is a 'core erythroid' enhancer, the sequence was synthesized and then put into the enhancer trap¹⁹ in one and four copies per plasmid and tested by transfection into HD3 cells as before. By this test, this core sequence alone is not sufficient to cause T-antigen accumulation over background, although it may be necessary for such activity; alternatively, it could be involved in some other unrelated aspect of cellular physiology.

We conclude that a 3' erythroid-specific enhancer element mediates the expression of the histone H5 gene by positively regulating its transcription in chicken erythroid cells. The function and location of this tissue-specific enhancer element (similar to the β-globin enhancer)^{2,3} suggests that enhancer-mediated transcriptional activation may be a recurrent theme in the orchestration of regulatory events governing erythroid gene expression during development.

We thank A. Ruiz-Carrillo for the original histone H5 cDNA clone, P. Boyer and J. Dodgson for genomic H5 clones, and Ok-Ryun Choi and James Gallarda for discussions. This work was supported by grants from the NIH, by a grant from the American Cancer Society and a Faculty Research Award (J.D.E.). It was also partially supported by an NIH National Research Services Award Traineeship (C.T.).

Received 9 June; accepted 1 July 1987.

1. Serfling, E., Jasin, M. & Schaffner, W. *Trends Genet.* **1**, 224–230 (1985).
2. Choi, O. R. & Engel, J. D. *Nature* **323**, 731–734 (1986).
3. Hesse, J. E., Nikol, J. M., Lieber, M. R. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4312–4316 (1986).
4. Neelin, J. M., Callahan, P. X., Lamb, D. C. & Murray, K. *Can. J. Biochem.* **42**, 1743–1752 (1964).
5. Weintraub, H. *Nucleic Acids Res.* **5**, 1179–1188 (1978).

6. Samarut, J. & Gazzolo, L. *Cell* **28**, 921-929 (1982).
7. Beug, H., Palmieri, S., Zentgraf, C. & Graf, T. *Cell* **28**, 907-919 (1982).
8. Beug, H., von Kirchbach, A., Döderlein, G., Conscience, J. & Graf, T. *Cell* **18**, 375-390 (1979).
9. Sortirov, N. & Johns, E. W. *Expl Cell Res.* **73**, 13-16 (1972).
10. Dalton, S., Coleman, J. R. & Wells, J. R. E. *Molec. cell. Biol.* **6**, 601-606 (1986).
11. Krieg, P. A., Robins, A. J., Gait, M. J., Titmas, R. C. & Wells, J. R. E. *Nucleic Acids Res.* **10**, 1495-1502 (1982).
12. Ruiz-Vazquez, R. & Ruiz-Carrillo, A. *Nucleic Acids Res.* **10**, 2093-2108 (1982).
13. Krieg, P. A., Robins, A. J., D'Andrea, R. & Wells, J. R. E. *Nucleic Acids Res.* **11**, 619-627 (1983).
14. Ruiz-Carrillo, A., Affolter, M. & Renaud, J. J. *molec. Biol.* **170**, 832-859 (1983).
15. Choi, O. R., Trainor, C., Graf, T., Beug, H. & Engel, J. D. *Molec. cell. Biol.* **6**, 1751-1759 (1986).
16. Palmieri, S., Beug, H. & Graf, T. *Virology* **123**, 296-311 (1982).
17. Beug, H., Döderlein, G., Freudenstein, C. & Graf, T. *J. cell. Physiol. suppl.* **1**, 195-207 (1982).
18. Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. & Howard, B. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6777-6781 (1982).
19. Weber, F., deVilliers, J. & Shaffner, W. *Cell* **36**, 983-992 (1984).
20. Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
22. Glisén, V., Crkvenjakov, R. & Byus, C. *Biochemistry* **13**, 2633-2637 (1974).
23. Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **7**, 1175-1193 (1979).

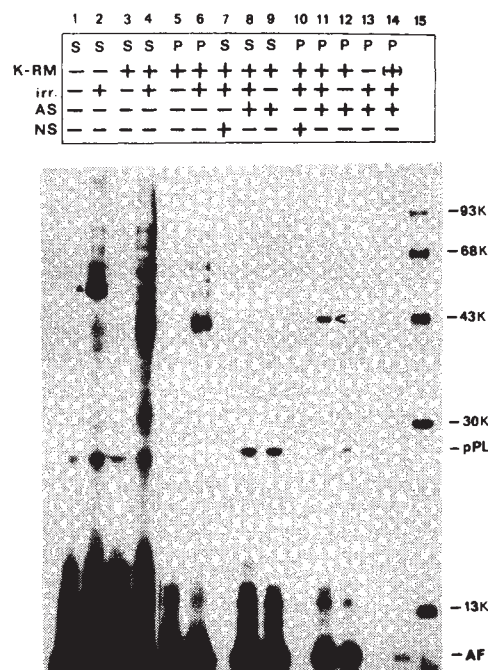


Fig. 1 Crosslinking of the AF of preprolactin with a signal sequence receptor (SSR) in the ER membrane. K-RM, high-salt washed membranes; irr., irradiation; AS, precipitation with anti-serum directed against sheep prolactin; NS, precipitation with non-immune serum. Triangle, AF*54K crosslinked product; pPL, preprolactin; AF, arrested fragment of preprolactin. Arrow head in lane 11 indicates the crosslinked product, consisting of the AF and an ER membrane protein, found in the carbonate extracted membrane pellet if the SRP-arrested translation complex was irradiated in the presence of microsomes.

Methods. Bovine pituitary mRNA was translated for 5 min in the presence of 40 nM lysyl-tRNA modified in the ϵ -position with 4-(3-trifluoromethyl-diazirino) benzoic acid (TDBA-Lys-tRNA) and 2 mCi ml⁻¹ ³⁵S-methionine in a wheat germ cell-free system supplemented with 80 nM SRP (isolated from dog pancreas²⁴) to produce an elongation-arrested complex^{1,6}. Cycloheximide (2 mM) was added and the incubation continued for 2 min. Where indicated, high-salt washed microsomes from dog pancreas (K-RM) were added (16 equivalents (eq.) per 50 μ l; see ref. 25) and the samples incubated for 4 min. Irradiation (irr.) as previously described⁶. To the sample shown in lane 14 membranes were added after irradiation. All samples were treated with 0.1 M Na₂CO₃ (pH 11.5-12) at 0°C and centrifuged. The supernatants (S) were either analysed directly (lanes 1-4) or after immunoprecipitation (lanes 7-9). The pellets (P) were washed twice with 0.1 M Na₂CO₃, dissolved in buffer A (0.1 M Tris/HCl pH 7.5, 1% SDS, 30 mM dithiothreitol (DTT)) and either analysed directly (lanes 5 and 6) or after immunoprecipitation (lanes 10-14). Immunoprecipitation as described⁶ with anti-serum (AS) (United States Biochem. Corp.) directed against sheep prolactin (lanes 8, 9, 11-14) or with non-immune serum (NS) (lanes 7 and 10). The proteins were separated on a 10-15% gradient polyacrylamide gel in SDS and visualized by fluorography of the diphenyloxazole (PPO)-impregnated gel. Lane 15: M, markers (lysozyme, 13K; carbonic anhydrase, 30K; ovalbumin, 43K; bovine serum albumin, 68K; phosphorylase A, 93K).

label components of the membrane that interact with the signal sequence.

The results of such an experiment are shown in Fig. 1. In agreement with previous results^{2,6}, irradiation of the arrested translation complex before the addition of membranes resulted in crosslinks between the arrested fragment of preprolactin (~8K) and the 54K polypeptide of SRP (referred to as AF*54K-product) (Fig. 1, lane 2, triangle). The formation of this product was greatly diminished if salt-extracted microsomal membranes were added before irradiation (compare lanes 2 and 4), indicating that the signal sequence had left the SRP. After irradiation,

A signal sequence receptor in the endoplasmic reticulum membrane

Martin Wiedmann, Teymuraz V. Kurzchalia,
Enno Hartmann & Tom A. Rapoport*

Zentralinstitut für Molekularbiologie der Akademie der
Wissenschaften der DDR, 1115 Berlin-Buch,
Robert-Roessle Strasse 10, DDR

Protein translocation across the endoplasmic reticulum (ER) membrane is triggered at several stages by information contained in the signal sequence. Initially, the signal sequence of a nascent secretory protein upon emergence from the ribosome is recognized by a polypeptide of relative molecular mass 54,000 (M_r 54K) which is part of the signal recognition particle (SRP)^{1,2,27}. Binding of SRP may induce a site-specific elongation arrest of translation *in vitro*^{1,3}. Attachment of the arrested translation complex to the ER membrane is mediated by the SRP-receptor (docking protein)^{3,4} and is accompanied by displacement of the SRP from both the ribosome⁵ and the signal sequence⁶. We have investigated the fate of the signal sequence following the disengagement of SRP and its receptor by a crosslinking approach^{2,6}. We report here that the signal sequence of nascent preprolactin, after its release from the SRP, interacts with a newly discovered component, a signal sequence receptor (SSR), which is an integral, glycosylated protein of the rough ER membrane (M_r ~35K).

The crosslinking approach adopted for this work allows us to identify components interacting with the signal sequence during the process of protein translocation. A photoreactive group is introduced into cell-free synthesized proteins by use of lysyl-tRNA modified in the ϵ -position of the amino acid; crosslinking to the interacting component is induced by irradiation^{2,6}. Bovine preprolactin was chosen as a model because it contains two lysine residues in positions 4 and 9 of its signal peptide of 25 residues, and no other lysine residue within the first 70 amino acids⁷. This is the size of the nascent polypeptide chain when SRP arrests elongation¹. Thus, any irradiation-induced crosslinked product can be attributed to an interaction of the signal sequence of the arrested fragment (AF). Two methionine residues present in the AF lend themselves to radioactive labelling.

A functional complex between the ER membrane and a ribosome carrying the AF was formed by incubation with dog pancreatic microsomes depleted of SRP by extraction with high salt (K-RM)^{8,9}. Chain elongation was prevented by prior addition of cycloheximide. Irradiation of the complex should affinity-

* To whom correspondence should be addressed.

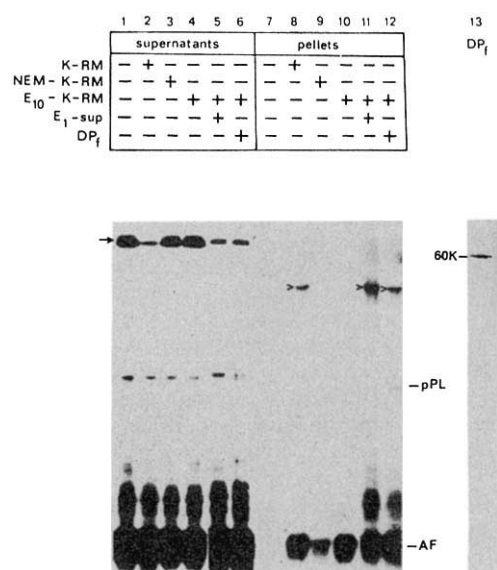


Fig. 2 Crosslinking of the AF with the SSR depends on the interaction of SRP with the SRP-receptor. Diminution of the AF*54K-crosslinked product (arrow) in the supernatant and appearance of the 43K crosslinked product (arrow head) in the carbonate-extracted membrane pellet on incubation of the SRP-arrested translation complex with intact (lanes 2 and 8) or reconstituted (lanes 5, 6 and 11, 12) microsomes. K-RM, microsomes washed with salt; NEM-K-RM, salt washed microsomes inactivated by treatment with *N*-ethylmaleimide (NEM); E₁₀-K-RM, salt-washed microsomes inactivated with elastase at 10 $\mu\text{g ml}^{-1}$; E₁-sup, supernatant from K-RM treated with 1 $\mu\text{g ml}^{-1}$ elastase; DP_r, cytoplasmic fragment of the SRP receptor.

Methods. Incubations as described in Fig. 1 legend. Samples 1-4, 7-10, translation in 50 μl ; samples 5, 6, 11, 12 in 100 μl . Supernatants and redissolved pellets of the carbonate extraction were immunoprecipitated with anti(prolactin) serum. Separation of proteins as in Fig. 1. Membranes were preincubated with soluble fractions for 30 min at 0 °C, centrifuged and washed before addition to the translation mixtures. K-RM, 16 eq. per 50 μl ; NEM-K-RM, the same, treated with 5 mM NEM for 30 min¹¹; E₁₀-K-RM, the same, treated at 0 °C for 60 min with 10 $\mu\text{g ml}^{-1}$ elastase and extracted with 0.5 M potassium acetate¹³; E₁-sup, supernatant obtained by treatment of ~100 eq. K-RM at 0 °C for 60 min with 1 $\mu\text{g ml}^{-1}$ elastase, addition of 1 mM phenylmethanesulphonyl fluoride (PMSF) and 0.5 M potassium acetate and removal of the membranes by sedimentation¹³; DP_r, cytoplasmic fragment of the SRP-receptor purified as described in ref. 14, ~1 μg of the 60K fragment per assay. Lane 13, silver stain of purified DP_r (about 0.2 μg applied); pPL, preprolactin; AF, arrested fragment of preprolactin.

the membranes were extracted with carbonate to strip off all luminal and peripheral membrane proteins¹⁰, thus yielding a supernatant fraction (S) containing these proteins and a pellet fraction (P) containing integral membrane proteins. A cross-linked product of $M_r \sim 43\text{K}$ resistant to alkaline extraction was found in the pellet of irradiated samples (lane 6). The 43K product was not seen in non-irradiated samples (compare lanes 5 and 6). The crosslinked product was immunoprecipitable with an antiserum directed against prolactin (lane 11, arrow head) but not with nonimmune serum (lane 10). Little of the 43K crosslinked product was found in the supernatant fraction (lanes 7-9). The crosslinked product thus behaves like an integral membrane protein, in contrast to the AF. As the AF of preprolactin has a M_r of ~8K, the integral membrane protein crosslinked to it may be assumed to have a M_r of ~35K.

Several controls were performed. The 43K band did not appear in the immunoprecipitates of the pellet fractions if the irradiation of the arrested translation complex was carried out in the absence of membranes (Fig. 1, lane 13) or if the membranes were added after crosslinking (lane 14; in this particular experi-

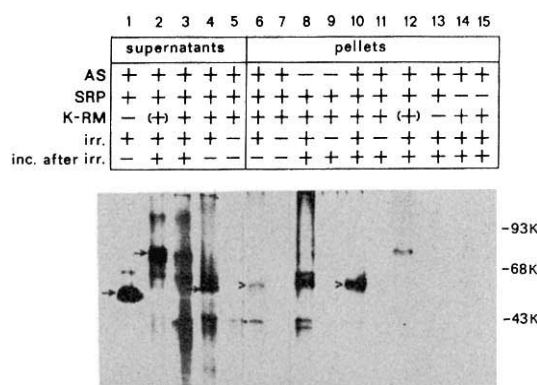


Fig. 3 Synthesis of completed preprolactin crosslinked to the SSR. Arrow head in lane 10 indicates the 60K crosslinked product of preprolactin and the SSR produced by irradiation of the arrested translation complex in the presence of microsomes and subsequent continuation of translation. Lane 6 shows the crosslinked products immediately after irradiation. Arrows: AF*54K product (lane 1) and pPL*54K product (lane 2). Dot in lane 6: 43K crosslinked product consisting of the AF and the SSR. Arrow heads in lanes 6 and 10: crosslinked product containing complete preprolactin and the SSR.

Methods. Translation in 50 μl with 40 nM SRP as in Fig. 1 legend. Sample 14 was incubated without SRP and sample 15 received 6 eq. K-RM instead. Irradiation in ice for 5 min as indicated (irr.) either before addition of K-RM (lanes 2 and 12) or after addition of K-RM (6 eq.) and incubation for 5 min at 0 °C (all other samples). Samples 1, 4-7 were analysed immediately, all others received 4 mM ⁷mGp, were warmed up to 26 °C, supplemented with 5 μl of 0.3 mM lysine, 3.6 mM GTP, 3.6 mM CTP, 20 mM DTT, 6 mM magnesium acetate and incubated for further 60 min (inc. after irr.). All samples were treated with carbonate and separated by centrifugation in supernatant and pellet. Samples 8 and 9 were analysed directly, all others immunoprecipitated with anti-prolactin serum. Separation of the proteins as in Fig. 1.

ment the amount of AF bound to the membranes was low since the incubation was carried out at 0 °C). The 43K-product was not seen if the modified lysyl-tRNA or SRP were omitted or if protein synthesis was inhibited by cycloheximide before addition of the mRNA or if the membranes were extracted with carbonate before irradiation. On the other hand, if after incubation with the arrested translation complex the membranes were treated with 2 M urea, 0.5 M potassium acetate or 25 mM EDTA and subsequently irradiated, the AF remained bound^{8,9} and the 43K band was produced (data not shown).

The appearance of the 43K crosslinked product depends on an interaction of the SRP with its membrane receptor (Fig. 2). Whereas untreated microsomes both diminished the yield of the AF*54K-product in the supernatant (compare Fig. 2, lanes 1 and 2, arrow) and gave rise to the 43K band in the pellet (compare Fig. 2, lanes 7 and 8, arrow head), microsomes pre-treated with *N*-ethylmaleimide (NEM) to inactivate the SRP-receptor¹¹ did not show these effects (Fig. 2, lanes 3, 9). Similar results were obtained with membranes that were treated with elastase or trypsin followed by extraction with high salt to remove the cytoplasmic domain of the SRP-receptor protein^{12,13}. SRP was not displaced from the signal sequence (Fig. 2, lane 4) and the 43K crosslinked band was not produced (Fig. 2, lane 10) in the presence of elastase-treated membranes (results for trypsin not shown). After complementation of the elastase-inactivated membranes with the cytoplasmic domain of the SRP-receptor (a 60K fragment) either in a crude state¹³ (Fig. 2, lanes 5, 11) or in a purified state¹⁴ (Fig. 2, lanes 6, 12; lane 13 shows the purity of the 60K fragment), both the release of the SRP from the signal sequence and the appearance of the 43K crosslinked product were restored.

To obtain an additional independent estimate of the size of the newly found signal sequence receptor (SSR), we made use

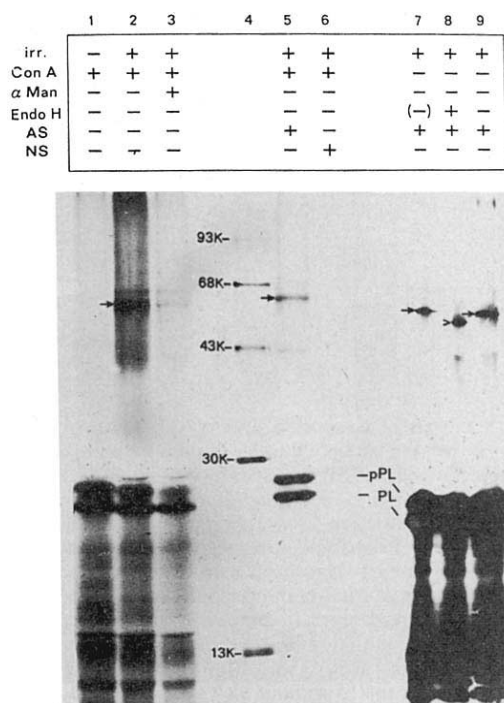


Fig. 4 The crosslinked product of preprolactin and the SSR is glycosylated. The 60K crosslinked product (arrows), consisting of completed preprolactin and the SSR, is bound to concanavalin A (Con A)-Sepharose in the absence of α -methylmannoside (α Man) (lanes 2 and 5) but not in its presence (lane 3). It is also sensitive to the action of endoglycosidase H (lane 8, arrow head). **Methods.** Translation in 100 μ l with 40 nM SRP as described in Fig. 1 legend. K-RM (12 eq.) were added at 0 °C and the incubation continued for 5 min. The samples were irradiated in ice for 5 min and subsequently warmed and incubated for a further 60 min as described in Fig. 3 legend. The membrane pellet resistant to alkaline extraction was dissolved in buffer A (see Fig. 1, legend). Samples for binding to Con A-Sepharose (samples 1-3) were diluted with Con A-buffer (70 mM Tris/HCl pH 7.5, 5 mM CaCl_2 , 1 mM MgCl_2 , 1.2% Triton X-100) and received 20 μ l Con A-Sepharose (Pharmacia). After incubation at room temperature for 4 h in the absence (samples 1, 2) or presence of 0.4 M α Man (sample 3) and washing with Con A-buffer, the proteins were eluted with buffer A at 95 °C for 5 min and equal aliquots were subjected to gel electrophoresis (lanes 1-3). In the case of sample 2, equivalent aliquots were subsequently immunoprecipitated with anti(prolactin) serum (AS) (lane 5) or nonimmune serum (NS) (lane 6). For treatment with endoglycosidase H (Endo H), the samples 7-9 of a separate experiment were first immunoprecipitated with a prolactin antiserum, then dissolved and analysed either directly (lane 9) or after incubation in the presence (lane 8) or absence (lane 7) of 0.1 units ml^{-1} endoglycosidase H (Boehringer Mannheim) at 37 °C for 8 h as described²⁶. All samples were precipitated with TCA, dissolved in SDS-sample buffer and subjected to gel electrophoresis in SDS. pPL, preprolactin; PL, prolactin.

of the fact that crosslinking of the signal sequence does not prevent further polypeptide chain elongation if translation is continued after irradiation (Fig. 3).

SRP-arrested polysomes were assembled and chilled on ice to prevent further polypeptide chain elongation. Irradiation in the absence of membranes yielded the AF*54K-product (Fig. 3, lane 1) that could be elongated subsequently to produce a product of $M_r \sim 80$ K containing completed preprolactin (~ 26 K) and the 54K polypeptide of SRP (referred to as the pPL*54K-product) (ref. 6; Fig. 3, lane 2). If salt-extracted microsomes were added before irradiation, the 43K crosslinked band was again seen in the immunoprecipitate of the carbonate pellet (Fig. 3, lane 6, dot) although it was less intense than in experiments in which elongation was inhibited by cycloheximide. In addition, a ~ 60 K-band was observed (Fig. 3, lane 6, arrow

head), probably owing to insufficient inhibition of translation by cooling. This 60K-band became a prominent product if the samples were warmed up after irradiation and translation was continued for 60 min (Fig. 3, lane 8), a time period amply sufficient to complete all polypeptide chains. The 60K-band was immunoprecipitable with an antiserum directed against prolactin (Fig. 3, lane 10, arrow head). It thus corresponds to the expected crosslinked product of completed preprolactin and the SSR (~ 26 K + ~ 35 K). The 60K band was not seen in nonirradiated controls (lanes 9 and 11) or in samples lacking membranes (Fig. 3, lane 13). The membranes had to be present before irradiation (compare lanes 12 and 10, Fig. 3). The addition of high-salt washed membranes (K-RM) in the absence of SRP did not yield the crosslinked product whether the membranes were present from the onset of translation (Fig. 3, lane 14) or were added after 5 min (Fig. 3, lane 15) at a time when the SRP-arrested complex had formed in samples containing SRP.

The 60K crosslinked product was found to be glycosylated (Fig. 4). It was bound by concanavalin A-Sepharose (Fig. 4, lane 2, arrow) and α -methylmannoside prevented the binding (Fig. 4, lane 3). The material eluted from concanavalin A was immunoprecipitable with an antiserum directed against prolactin (Fig. 4, lane 5, arrow) but not with a nonimmune serum (lane 6). As bovine (pre)prolactin itself is not glycosylated, the crosslinked membrane component must be responsible for the interaction with the lectin. This conclusion is supported by the sensitivity of the 60K crosslinked product to the action of endoglycosidase H which produced a band of higher mobility (Fig. 4, lane 8, arrow head) compared with the untreated (lane 9, arrow) or mock-treated samples (lane 7, arrow).

Our results show that an integral, glycosylated membrane protein (M_r 34-35K) that spans the phospholipid bilayer acts as a receptor for the signal sequence of preprolactin after binding of the SRP-arrested translation complex to the rough ER membrane. The existence of a signal sequence receptor (SSR) has been suggested before on the basis of indirect evidence^{8,9,15-18}. Recently, by use of a highly artificial ligand, a synthetic 'consensus' signal peptide carrying a photoreactive group, a cross-linked membrane protein of $M_r \sim 45$ K was found by others¹⁹. The discovery of the SSR argues against an exclusive interaction of the signal sequence with the lipid and supports the view that the ER translocation system consists of several protein components.

What might be the role of SSR in protein translocation? SSR could mediate the interaction of the signal sequence with the lipid, it could induce or open a channel through which the polypeptide traverses the membrane²⁰ or it could even be part of such a channel.

As two entities have been identified that recognize signal sequences (SRP and SSR) the question of what their specific contributions to protein translocation may be arises. We presume that the function of SRP is to prevent folding of the polypeptide into an unfavourable conformation before translocation, whereas the basic mechanism of targeting may be the interaction of a signal sequence with SSR. Indeed, some proteins seem to be transported post-translationally across the membrane in an SRP-independent manner^{21,22}. In these cases the conformation of the polypeptide, even if completed, may still allow an interaction with the SSR. A receptor similar to SSR may be responsible for the post-translational import of proteins into mitochondria²³ or other cell organelles.

We thank Professor H. Bielka for continuous support and Professor S. M. Rapoport and Dr H. Welfle for critical reading of the manuscript. We are grateful to Drs A. S. Girshovich and E. S. Bochkareva, Institute of Protein Research Poustchino (USSR) for supplying TDBA.

Received 20 May; accepted 24 June 1987.

1. Walter, P. & Blobel, G. *J. Cell Biol.* **91**, 557-561 (1981).
2. Kurzchalia, T. V., Wiedmann, M., Girshovich, A. S., Bochkareva, E. S., Bielka, H. & Rapoport, T. A. *Nature* **320**, 634-636 (1986).

3. Meyer, D. I., Krause, E. & Dobberstein, B. *Nature* **297**, 503-508 (1982).
4. Gilmore, R., Walter, P. & Blobel, G. *J. Cell Biol.* **96**, 470-477 (1982).
5. Gilmore, R. & Blobel, G. *Cell* **35**, 677-685 (1983).
6. Wiedmann, M., Kurzchalia, T. V., Bielka, H. & Rapoport, T. A. *J. Cell Biol.* **104**, 201-209 (1987).
7. Sasavage, N. L., Nilson, J. H., Horowitz, S. & Rottman, F. M. *J. biol. Chem.* **257**, 678-681 (1982).
8. Gilmore, R. & Blobel, G. *Cell* **42**, 497-505 (1985).
9. Conolly, T. & Gilmore, R. *J. Cell Biol.* **103**, 2253-2261 (1986).
10. Fujiki, Y., Hubbard, A. L., Fowler, S. & Lazarow, P. B. *J. Cell Biol.* **93**, 97-102 (1982).
11. Gilmore, R., Blobel, G. & Walter, P. *J. Cell Biol.* **95**, 463-469 (1982).
12. Walter, P., Jackson, R. J., Marcus, M. M., Lingappa, V. R. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 1795-1799 (1979).
13. Meyer, D. I. & Dobberstein, B. *J. Cell Biol.* **87**, 503-508 (1980).
14. Meyer, D. I. & Dobberstein, B. *Meth. Enzym.* **96**, 692-700 (1983).
15. Prehn, S., Tsamaloukas, A. & Rapoport, T. A. *Eur. J. Biochem.* **107**, 185-195 (1980).
16. Prehn, S., Nuernberg, P. & Rapoport, T. A. *FEBS Lett.* **123**, 79-84 (1981).
17. Majzoub, J. A. *et al. J. Biol. Chem.* **255**, 11478-11483 (1980).
18. Austen, B. M. & Ridd, D. H. *Biochem. Soc. Trans.* **11**, 160-161 (1983).
19. Robinson, A., Kaderbhai, M. A. & Austen, B. *Biochem. J.* **242**, 767-777 (1987).
20. Blobel, G. & Dobberstein, B. *J. Cell Biol.* **67**, 852-862 (1975).
21. Mueckler, M. & Lodish, H. F. *Cell* **44**, 629-637 (1986).
22. Zimmermann, R. & Mollay, C. *J. biol. Chem.* **261**, 12889-12895 (1986).
23. Zwizinski, C., Schleyer, M. & Neupert, W. *J. biol. Chem.* **259**, 7850-7856 (1984).
24. Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 682-691 (1983).
25. Walter, P., Ibrahim, I. & Blobel, G. *J. Cell Biol.* **91**, 545-550 (1981).
26. Anderson, D. J. & Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **78**, 5598-5602 (1981).
27. Krieg, U., Walter, P. & Johnson, A. E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8604-8608 (1986).

Sulphur radicals formed by cutting α -keratin

Harish Chandra & Martyn C. R. Symons

Department of Chemistry, The University, Leicester LE1 7RH, UK

Finger nails are composed largely of α -keratin, a protein in which three right-handed α -helical peptide chains are twisted into a left-handed coil strengthened by disulphide cross links. We have found that the act of cutting nails generates trapped radicals, giving intense electron spin resonance (ESR) signals characteristic of sulphur centred radicals. The nature of the primary radical and the possible modes of generation are discussed. These radical reactions should be considered when using human nail parings to estimate accidental exposure to ionizing radiation.

Radicals in solids can be generated by grinding, generally at low temperatures. These radicals are detected by ESR spectroscopy, but the mechanism of their formation is not clear. No examples in which trapped radicals have been formed by the related process of cutting have been reported. Here we investigate the nature of the ESR signal generated by cutting human nails at room temperature. We find that radicals are not detectable before or after the initial cut, but as the extent of cutting is increased, so the intensity of the ESR spectrum increases. We have not reached a saturation level for nail particles in the 0.1-1.0 mm size range. The signals begin to decay at $\sim 45^\circ\text{C}$. Typical spectra, at low and high microwave powers (0.5 and 5.0 mW) are shown in Fig. 1. These were recorded at 77 K, but were still reasonably well-defined at room temperature. The main species, having a single proton splitting of $A(^1\text{H}) \sim 8.5$ G and peak g-values at 2.061, 2.025, and 2.000, is clearly fingerprinted as a di-sulphur species sometimes formulated as RSS^{1-4} and sometimes as $\text{RS}\dot{\text{S}}\text{R}^{5,6}$. Following the realization that this species is not an RS radical^{1-3,7}, these alternative formulations have been extensively discussed, the current consensus being in favour of the RSS formulation⁸. We need to consider why this specific species is formed in high abundance.

It is the disulphide cross-links in α -keratin that are implicated in the present study, and we would expect that mechanical stress would break the weak S-S bonds in favour of the stronger C-S bonds. The former generates two RS radicals, and if these are prevented from recombining, they may become the source of the ESR signals. Such radicals will not be detectable directly by ESR because their features are expected to be very broad⁷, but they have a high affinity for sulphur, and reactions (1) and

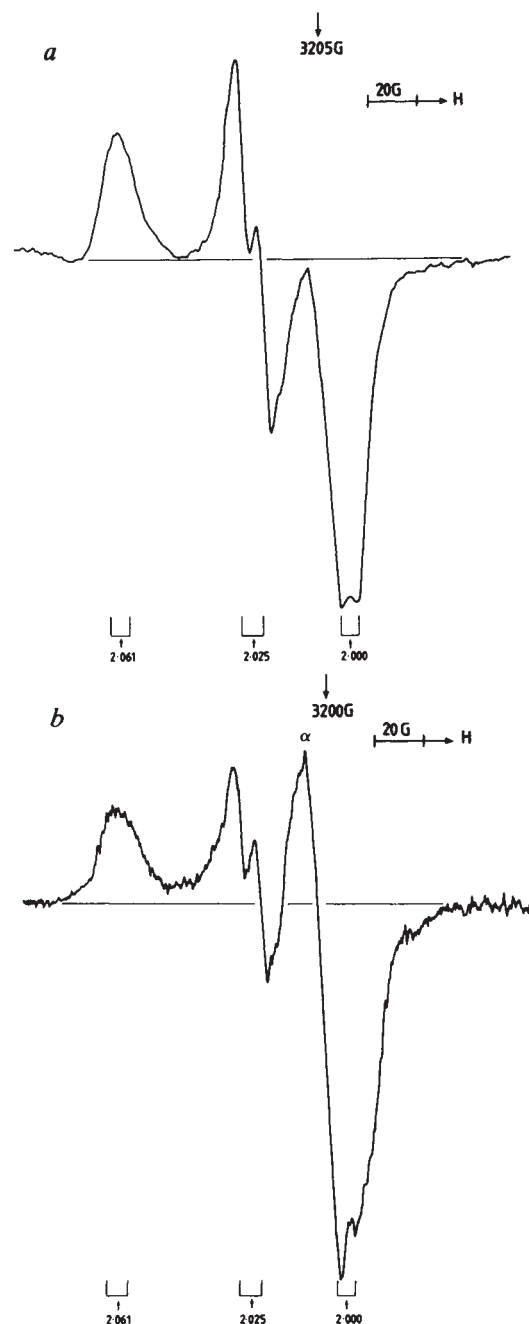
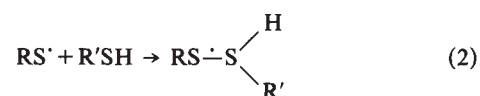
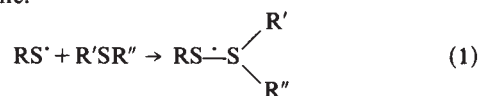


Fig. 1 First derivative X-band ESR spectra for small pieces of finger nails, taken at 77 K. There are three g-features shown, split into doublets, which are assigned to $\text{RS}\dot{\text{S}}\text{R}''$ or $\text{RS}\dot{\text{S}}(\text{H})\text{R}'$ radicals, together with a second species thought to be $(\text{RS}\dot{\text{S}}\text{R})^-$ radical anions. α , At high power (5 mW) and b , at low power (0.5 mW), showing the α peak secondary feature more clearly.

(2) should be facile.



We suggest that these are the species responsible for the major ESR signals, the unique proton splitting being associated with one β -proton of the R-group. We cannot rule out the $\text{RSS}\cdot$ structure for these radicals, but consider that their formation is chemically less probable.

The second species (designated ' α ' in Fig. 1b), which is seen more clearly at lower microwave power, is poorly defined and in low abundance, but it occurs on the low field side of $g = 2.0023$ (the free-spin value), with a maximum value of g of ~ 2.018 . This is therefore also a sulphur species, the spread in g -values being close to those reported for the radical anion, $(RS^{\cdot-}SR)^{\cdot-}$. These anions cannot be formed by bond homolysis or in any direct manner. It is possible that they are formed by a 'triboelectric' effect or by a partial deprotonation of radicals produced in the reaction shown in equation (2). This would account for the constant presence of both species, and is reasonable because the pK_a value of these radicals is thought to be close to the neutral point⁹.

We conclude that cutting α -keratin results in radical formation, the products, as judged by ESR spectroscopy, being remarkably specific. These results should be considered in measurements of free radical production resulting from radiation damage.

This work was supported by the National Radiation Protection Board, Harwell.

Received 27 May; accepted 7 July 1987.

1. Hadley, J. H. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3486-3492 (1975).
2. Hadley, J. H. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.* **71**, 3106-3112 (1974).
3. Hadley, J. H. & Gordy, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 216-223 (1977).
4. Henriksen, T. *J. chem. Phys.* **37**, 2189-2198 (1962).
5. Nelson, D. J., Petersen, R. L. & Symons, M. C. R. *J. chem. Soc. Perkin Trans. II*, 2005-2015 (1977).
6. Nelson, D. J., Petersen, R. L. & Symons, M. C. R. *J. chem. Soc. Perkin Trans. II*, 225-231 (1978).
7. Symons, M. C. R. *J. chem. Soc. Perkin Trans. II*, 1618-1622 (1974).
8. Symons, M. C. R. in *Oxygenated Sulphur Radicals in Chemistry and Medicine* (eds Breccia, A., Rodgers, M. A. J. & Semeraro, G.) 71-88 (Lo Scarabeo, Bologna, 1986).
9. Hoffman, M. Z. & Hayon, E. *J. Am. chem. Soc.* **94**, 7950-7963 (1972).

The accessible surface area and stability of oligomeric proteins

Susan Miller*, Arthur M. Lesk†‡, Joël Janin§ & Cyrus Chothia*†

*Christopher Ingold Laboratories, University College London, 20 Gordon Street, London WC1H 0AJ, UK

†Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

‡Biocomputing Programme, European Molecular Biology Laboratory, Meyerhofstrasse 1, 6900 Heidelberg, FRG

§Laboratoire de Biologie Physicochimique, Université Paris-Sud, Bât. 433, 91405 Orsay, France

Protein structures are stabilized by hydrophobic and van der Waals forces, and by hydrogen bonds. The relation between these thermodynamic quantities and the actual three-dimensional structure of proteins can not be calculated precisely. However, certain empirical relations have been discovered. Hydrophobic energy is gained by the reduction of surface in contact with water¹. For monomeric proteins, the area of the surface accessible to solvent, and of that buried in the interior, is a simple function of molecular weight. Proteins with different shapes and secondary structures, but of the same molecular weight, have the same accessible surface area²⁻⁵. It has been argued that there is no similar relationship for large oligomeric proteins⁶. In this paper we show that the surface areas of oligomeric proteins, and the areas of the surface buried within them, are directly related to relative molecular mass. Although oligomers of the same molecular weight bury the same amounts of surface, the proportions buried within and between subunits vary. This has important implications for the role of subunit interfaces in the stability and activity of oligomeric proteins.

Calculations were carried out on the 23 oligomers listed in Table 1. All these structures are well determined as judged by the resolution and R factor of the structure determination, the

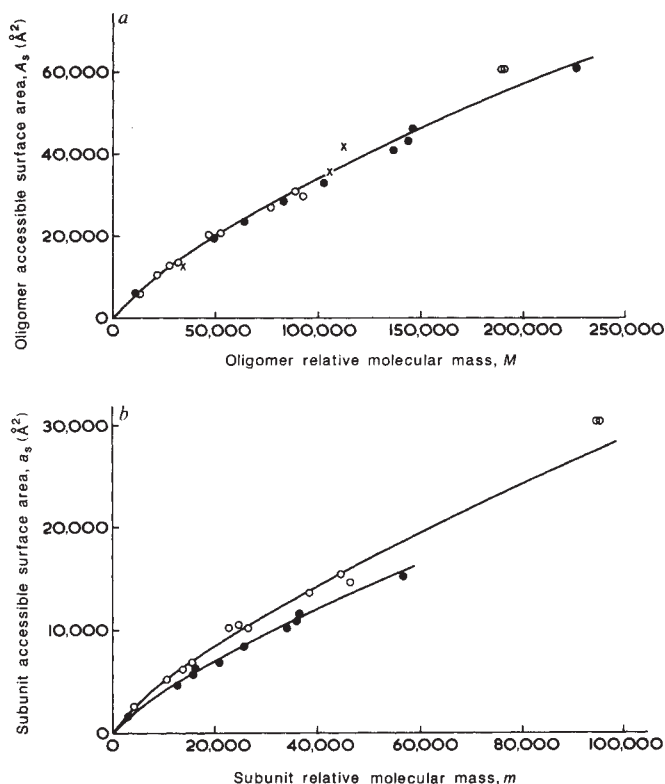


Fig. 1 a, The accessible surface areas of the oligomeric proteins (A_s) in Table 1 plotted against relative molecular mass (M). The data for their subunits are given in Table 2. The values for dimeric proteins are indicated by $\circ$, those for tetramers by $\bullet$, and those for the hexamers and octamer by $\times$. The curve in the drawing is given by the equation $A_s = 5.3M^{0.76}$ (see text). b, Accessible surface area of the subunits (a_s) of dimers ($\circ$) and tetramers ($\bullet$) are plotted against their relative molecular mass (m). The two curves in the drawing are given by the equations $a_s = 4.5m^{0.76}$ and $a_s = 3.8m^{0.76}$ (see text).

torsion angles of the polypeptide chain and the hydrogen bonding of buried polar groups. Atomic coordinates for these proteins were obtained from the Protein Data Bank⁷ or were given to us by their authors. Table 1 gives the references to the crystallographic analyses that determined the structures.

The accessible surface area⁸ of individual protein atoms were calculated using the Shrake and Rupley⁹ algorithm in the manner described previously⁵. For each protein, we determined the accessible surface area of the monomer in isolation and of the monomer as part of the oligomer (Table 2). The area of the monomer surface buried in the subunit interfaces is the difference between these two numbers.

In Fig. 1a the accessible surface area of the oligomers is plotted against their molecular mass. This graph shows a direct relationship between these two quantities. Oligomers with different shapes, secondary structures and number of subunits, but with the same molecular mass (M), have very similar accessible surface area (A_s). If only the high molecular weight proteins are considered there is a linear relation between A_s and M . Over the whole range of molecular weights, the equation

$$A_s = 5.3M^{0.76} \quad (1)$$

gives A_s values for 17 of the proteins that deviate by no more than 5% from the observed values. For the other five proteins the deviations are 8-13%.

It might be thought that, because equations of the form $A_s = kM^{2/3}$ would apply to any set of solid bodies having the same shapes and densities, equation (1) is simply the geometrical consequence of polypeptide chains folding to form globular structures. This, however, is not the origin of equation (1). The

Table 1 Protein structures used in this work

Protein	Residues in monomer	Data bank file number* and structure reference	
Dimers			
Avian pancreatic polypeptide	36	1PPT	14
Subtilisin inhibitor	107	2SSI	15
Cytochrome <i>c'</i>	128	2CCY	16
Superoxide dismutase	152	2SOD	17
Fab KOL	216/229	1FB4	18
Triose phosphate isomerase	247	1TIM	19
Alcohol dehydrogenase	374	—	20
Aspartate aminotransferase	401	—	21
Citrate synthase	432	3CTS	22
Phosphorylase <i>a</i>	842	—	23
Phosphorylase <i>b</i>	842	—	†
Tetramers			
Melittin	26	1MLT	24
Prealbumin	127	2PAB	25
Haemoglobin	141/146	3HHB	26
Glutathione peroxidase	198	1GPI	27
Concanavalin A	237	2CNA	28
Phosphofructokinase	319	—	29
Lactate dehydrogenase	329	—	30
Glyceraldehyde-3-phosphate dehydrogenase	334	—	31
Catalase	506	7CAT	32
Hexamers			
Insulin	51	1INS	33
Phycocyanin	162/172	—	34
Octomer			
Haemerythrin	113	1HMQ	35

* Atomic coordinates were taken from the named file of the protein structure databank⁷; except for those proteins marked (—), coordinates for these structures were given to us by their authors.

† The 1.9 Å resolution atomic structure of phosphorylase *b* was determined by K. R. Acharya, D. I. Stuart and L. N. Johnson (personal communication).

proteins studied here differ greatly in shape: some of the dimers and tetramers are 'ellipsoids' or 'dumb-bells'; in other tetramers the four subunits extend in tetrahedral directions away from the interfaces (see the references in Table 1). Thus in these proteins the polypeptide chains fold into shapes that can be very different but whose surface areas are related to their molecular masses by equation (1).

A recent analysis of 37 well determined monomeric proteins showed that the expression:

$$A_S = 6.3M^{0.73} \quad (2)$$

gives A_S values that on average are within 4% of those observed⁵. (The parameters in this equation differ slightly from those previously obtained with less well determined structures^{3,4}.) The molecular mass range of the monomeric proteins used to derive equation (2) is 4,000 to 35,000 (ref. 5). In this range the A_S values given by equation (2) are 7% to 13% lower than those given by equation (1). The significance of this small difference is unclear.

Equation (1) implies that subunits of the same molecular weight have lower accessible surface areas in tetramers than they have in dimers. If m is the molecular weight of a subunit, the accessible surface area, a_s , that we would expect it to have if it is part of a dimer is:

$$a_s = \frac{A_S}{2} = \frac{5.3}{2} (2m)^{0.76}$$

$$a_s = 4.5 m^{0.76}$$

If the subunit is part of a tetramer we would expect:

$$a_s = \frac{A_S}{4} = \frac{5.3}{4} (4m)^{0.76}$$

$$a_s = 3.8 m^{0.76}$$

Figure 1b shows that the subunits in the dimers and tetramers have observed accessible surface areas close to the expected values.

The total accessible surface area of unfolded proteins (A_T) can be estimated by calculating the accessible surface area of the polypeptide in an extended conformation. The expression:

$$A_T = 1.48M \quad (3)$$

fits calculated values to within 1% on average⁵. Because the surface area that remains accessible when oligomers fold is a function of molecular weight, the surface buried within oligomers, $A_T - A_S$, is also a function of molecular weight and its value can be calculated from equations (1) and (3):

$$A_T - A_S = 1.48M - 5.3M^{0.76}$$

The surface buried within an oligomer is the sum of that buried within the subunits and of that buried in the interfaces between them. For the oligomers studied here the size of the surface buried between the subunits varies widely, from 1,360 Å² in superoxide dismutase to 42,300 Å² in catalase (Table 2). Of particular interest is the variation in the relative contributions the interfaces make to the total buried surface. The area buried in the interface represents between 3% and 40% of the total buried surface $A_T - A_S$. Thus, the contribution of the surfaces buried within and between subunits varies greatly, even though the total buried surface is essentially the same for proteins of a given molecular mass.

In oligomers with small interfaces the isolated subunits have accessible surface areas similar to those found in monomeric proteins of the same molecular mass. For example, the values for the isolated subunits of superoxide dismutase, prealbumin, haemoglobin, glutathione peroxidase, insulin and haemerythrin (Table 2) are within 4% of the values given by equation (2). Conversely, in the oligomers with large interfaces, the isolated

Table 2 Accessible surface area of the monomers in oligomeric proteins ($\text{\AA}^2$)

Protein	Subunit molecular mass	Accessible surface area of monomer in		
		oligomer	isolation	interface
Dimers				
Avian pancreatic polypeptide	4,250	2,630	3,320	690
ROP protein	6,650	3,050	4,390	1,340
Subtilisin inhibitor	10,910	5,470	6,220	750
Cytochrome <i>c</i> '	13,910	6,340	7,150	810
Superoxide dismutase	15,550	6,760	7,440	680
Fab KOL VL-CL	22,880	10,220	12,050	1,830
VH-CH1	24,300	10,500	12,280	1,780
Triose phosphate isomerase	26,510	10,130	11,720	1,580
Alcohol dehydrogenase	40,590	13,610	15,240	1,630
Aspartate aminotransferase	45,150	15,440	18,590	3,150
Citrate synthase	47,930	14,540	19,420	4,880
Phosphorylase <i>a</i> *	95,300	30,370	33,820	3,450
Phosphorylase <i>b</i> *	95,350	30,380	32,690	2,310
Tetramers				
Melittin	2,830	1,560	2,630	1,070
Prealbumin	12,450	4,890	6,400	1,510
Haemoglobin α	15,780	5,700	7,420	1,720
β	16,390	6,340	7,840	1,500
Glutathione peroxidase	20,790	7,090	8,680	1,590
Concanavalin A	25,540	8,300	10,880	2,580
Phosphofructokinase	35,310	10,150	13,750	3,600
Lactate dehydrogenase	36,490	11,700	17,240	5,540
Glyceraldehyde-3-phosphate dehydrogenase	36,560	10,810	14,610	3,800
Catalase	57,940	15,220	25,780	10,570
Hexamers				
Insulin subunit 1	5,740	2,080	3,510	1,430
subunit 2	5,740	2,250	3,680	1,430
Phycocyanin α	18,010	6,600	8,800	2,200
β	19,380	7,200	9,600	2,400
Octomer				
Haemerythrin	13,300	4,480	6,190	1,7120

* In phosphorylase *a* and *b* different regions of the protein are disordered: residues 314 to 324 and 835 to 840 in *a* and residues 1 to 18 in *b*. These residues are not included in the accessible surface areas or in the subunit molecular weights given above.

† The subunit molecular weights given in this table do not include any parts of the protein for which the structure was not determined. They do include any co-factor present in the structure.

subunits have accessible surface areas much larger than those of monomeric proteins of the same size. For example, the values for the isolated subunits of melittin and lactate dehydrogenase are 26–28% larger than those given by equation 2. In these structures we would expect the quaternary interactions to make an important contribution to the conformational stability of the subunits, as is also suggested by experimental studies of their folding¹⁰.

The structural features of oligomeric proteins reported here can be related to their thermodynamic and functional properties. Free energies of protein folding are small differences between large terms favouring either the unfolded or folded state^{11,12}, particularly conformational entropy of the polypeptide chain and hydrophobic free energy. Conformational entropy favours

the unfolded state and depends upon the length of the polypeptide chain, while hydrophobicity depends upon contacts with the solvent and favours the folded state^{1,13}. The balance between these two terms implies that solvent accessible surface should be correlated with molecular weight. The correlation has been established in monomeric proteins^{2–5}. Here it is established for oligomeric proteins which have the added complexity of subunit interfaces.

We thank John Cresswell for the figure; Drs H. Eklund, P. R. Evans, R. J. Fletterick, R. Huber, J. N. Jansonius, L. Johnson, T. A. Jones, M. G. Rossmann, T. Schirmer and A. J. Wonacott for the atomic coordinates of protein structures; and the Royal Society, the National Institute of General Medical Sciences and the National Science Foundation for support.

Received 22 April; accepted 13 July 1987.

- Kauzmann, W. *Adv. Protein Chem.* **14**, 1–63 (1959).
- Chothia, C. *Nature* **254**, 304–308 (1975).
- Janin, J. *J. molec. Biol.* **105**, 13–14 (1976).
- Teller, D. *Nature* **260**, 729–731 (1976).
- Miller, S., Janin, J., Lesk, A. M. & Chothia, C. *J. molec. Biol.* **196**, 641–656 (1987).
- Sprang, S., Yang, D. & Fletterick, R. J. *Nature* **280**, 333–335 (1979).
- Bernstein, F. C. *et al. J. molec. Biol.* **112**, 535–542 (1977).
- Lee, B. & Richards, F. M. *J. molec. Biol.* **55**, 379–400 (1971).
- Shrake, A. & Rupley, J. A. *J. molec. Biol.* **79**, 351–371 (1973).
- Jaenicke, R. *Angew. Chem.* **23**, 395–413 (1984).
- Privalov, P. L. *Adv. Protein Chem.* **33**, 167–241 (1979).
- Creighton, T. *Proteins* (Freeman, New York, 1983).
- Baldwin, R. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 8069–8072 (1986).
- Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4175–4179 (1981).
- Mitsui, Y., Satow, Y., Watanabe, Y. & Iitaka, Y. *J. molec. Biol.* **131**, 697–724 (1979).
- Finzel, B. C., Weber, P. C., Hardman, K. D. & Salemme, F. R. *J. molec. Biol.* **186**, 627–643 (1985).
- Tainer, J. A., Getzoff, E. D., Beem, K. M., Richardson, D. C. & Richardson, J. S. *J. molec. Biol.* **160**, 181–217 (1982).

- Marquart, M., Deisenhofer, J. & Huber, R. *J. molec. Biol.* **141**, 369–391 (1980).
- Banner, D. W. *et al. Nature* **255**, 609–614 (1975).
- Eklund, H. *et al. J. molec. Biol.* **146**, 561–587 (1981).
- Ford, G. C., Eichele, G. & Jansonius, J. N. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2559–2563 (1980).
- Remington, S., Wiegand, G. & Huber, R. *J. molec. Biol.* **158**, 111–152 (1982).
- Sprang, S. & Fletterick, R. J. *J. molec. Biol.* **131**, 523–551 (1979).
- Terwilliger, T. C. & Eisenberg, D. *J. biol. Chem.* **257**, 6016–6022 (1982).
- Blake, C. C. F., Geisow, M. J., Oatley, S. J., Rerat, B. & Rerat, C. *J. molec. Biol.* **121**, 339–356 (1978).
- Fermi, G., Perutz, M. F., Shaanan, B. & Fourme, R. *J. molec. Biol.* **175**, 159–174 (1984).
- Epp, O., Ladenstein, R. & Wendel, A. *Eur. J. Biochem.* **133**, 51–69 (1983).
- Reeke, G. N., Becker, J. W. & Edelman, G. M. *J. biol. Chem.* **250**, 1525–1547 (1975).
- Evans, P. R. & Hudson, P. J. *Nature* **279**, 500–504 (1979).
- White, J. L. *et al. J. molec. Biol.* **102**, 759–779 (1976).
- Biesecker, G., Harris, J. I., Thierry, J.-C. & Wonacott, A. J. *Nature* **266**, 328–333 (1977).
- Murthy, M. R. N., Reid, T. J., Scignano, A., Tanaka, N. & Rossmann, M. G. *J. molec. Biol.* **152**, 465–499 (1981).
- Blundell, T. L., Dodson, G. G., Hodgkin, D. C. & Mercola, D. *Adv. Protein Chem.* **26**, 279–402 (1972).
- Schirmer, T., Bode, W., Huber, R., Sidler, W. & Zuber, H. *J. molec. Biol.* **184**, 257–277 (1985).
- Stenkamp, R. E., Sieker, L. C. & Jensen, L. A. *Acta crystallogr.* **B39**, 697–703 (1983).

FT-IR microspectrometry advances

J.M. Kwiatkoski and J.A. Reffner

Fourier transform infrared (FT-IR) microspectrometry has created a renaissance in infrared analysis, with applications from cosmic dust to the analysis of plastic laminates.

THERE is a natural synergism between light microscopy and Fourier transform-infrared (FT-IR) spectroscopy when both are combined. When a microscopist examines a sample, a major goal is to identify the various constituents of the sample. The geometry, colour and optical properties are valuable identification features. Add to this the ability to obtain molecular spectra of each observed phase and a new level of analysis is achieved.

Microscopy differentiates IR microspectrometry from conventional IR microsampling techniques. With the technique, it is possible to see and define exactly the sampling area for spectroscopic analysis, and analyse physical mixtures without prior separation of constituents or deconvolution of complex spectra. Most microsampling methods for IR spectroscopy do not permit visual examination of the sample, they only reduce the amount of sample required for analysis. With microsampling the resultant spectrum is the sum of the absorptions from all components. IR microspectrometry yields the spectra of individual phases.

The first commercial infrared microscope attachment for IR spectrometry, offered by Perkin-Elmer in 1953, was designed by Coates, Offner and Siegler¹. The objective and condenser of this microscope were Schwarzschild-type² reflecting lenses. This microscope was produced for Perkin-Elmer Models 12, 112 and 13 single-beam IR spectrometers.

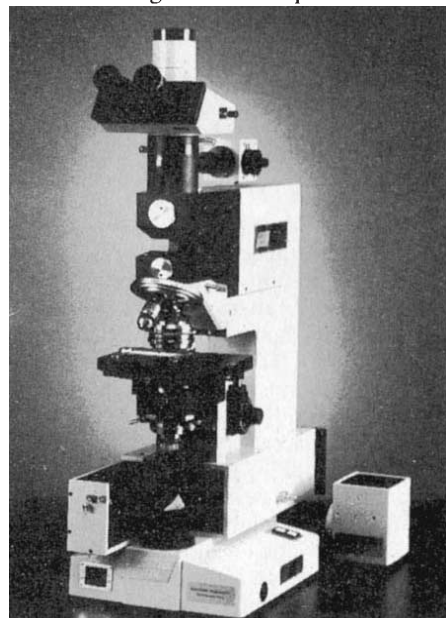


Fig. 1 The Spectra-Tech IR-PLAN microscope for FT-IR microspectrometry.

In the earlier work of Thompson and Gore^{3,4}, the microscope was placed between the radiation source and the monochromator. Later, the Coates design was improved by placing the microscope between the monochromator and the radiation detector to avoid heat and radiation damage to sensitive samples. While this microscope, the Perkin-Elmer Model 85, produced reasonable spectra of microscopic samples, it was not a commercial success and production was limited.

The development of the FT-IR spectrometer has caused a renaissance in infrared spectroscopy and has made IR microspectrometry a viable technology. General-purpose microscopes that combine the high quality visual imaging of research light microscopes with reflecting optics for IR microspectrometry are now available from several manufacturers. This microscope (Fig. 1) has both transmission and reflection illumination, and meets the Koehler illumination criterion for imaging. IR spectra can be collected by transmission through the sample or by reflection from the surface of the sample. The visible light path is coaxial with the IR spectrometer's light path to ensure that the region of the sample selected visually is the same as that being analysed. Remote, variable apertures are placed at conjugate image planes to define the sample area and to reduce stray light for spectral measurements; a technique known as redundant aperturing.

Selected applications

From the high technology of modern photographic materials to the lowly trash bag, plastic laminates are finding increased use in today's world. Analysis and quality control of these laminates dictate the need to ensure that the correct layers are present in the proper sequence and that there are no defects or contaminants. Modern laminates can have several layers that may be bonded together with adhesives. Chemical separation or physically peeling layers can be complex, incomplete and impractical procedures for analysis. FT-IR microspectrometry provides a direct solution to this analytical problem. After the separation of laminate cross-sections by conventional microtomy, the IR spectral analysis of each layer's infrared spectrum is recorded after it is optically isolated with the dual remote variable apertures of the microscope (Fig. 2). This reduces an intractable analytical separation problem to a simple direct examination — essentially a one-step operation.

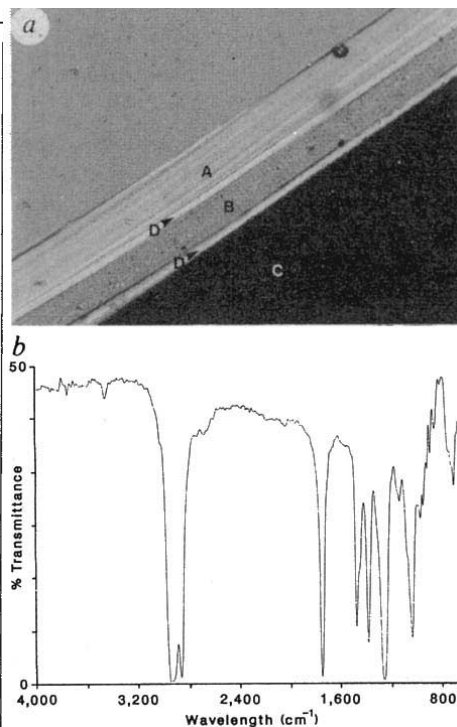


Fig. 2 *a*, Layers of a plastic cup. A Corresponds to polyethylene, B to polyvinylidene chloride, and C to polystyrene. *b*, Spectrum of the adhesive layer D from polymer laminate, characterized as ethylene vinylacetate.

Many natural and artificially produced materials exist as mixtures of solids. Traditional methods for characterizing such mixtures rely on bulk analysis, where all components are analysed as one single entity, or on physical separation of the sample. A physical separation is usually necessary if the identity of the individual components is required. Most samples require wet chemical or chromatographic methods to produce the individual components in a state of purity adequate for accurate analytical results.

Under a light microscope it is often possible to separate this type of sample into its individual components as a function of crystallinity or morphology. Even a seemingly uniform crystalline mixture can be separated in this manner into crystals of different shapes or even different birefringence, if cross-polarized light is used. With FT-IR microspectrometry, once a mixture is separated using this procedure — a task that may often take only a few seconds — an FT-IR analysis can be performed on the individual crystals. The scaling down of the physical separation step to microscopic dimensions can make many tedious procedures, such as fractional crystallization, useful tools.

The applications of FT-IR microspectrometry include drug mixtures, fibres and polymers, minerals, toxic wastes, consumer products, cosmetics, and food products⁵, to name but a few. Pharmaceutical products are a large area of interest because the FT-IR microscope may be used for the study of "pure" drugs, as well as drug preparations containing sugars and fillers.

The technique can also be used to study pure drugs that exist in polymeric forms. Often the morphology of a drug is critical to its physiological behaviour, because it may have reduced or no activity in certain crystalline forms. For some classes of drugs it is important for the analytical chemist to be aware of the drug's alternate morphologies, because the different forms may have very different infrared spectra, and this can lead to incorrect or misleading results when attempting to match spectra with a data base.

The range of FT-IR microspectrometry analyses extends from the common debris of forensic trace evidence⁶ to the far reaches of interplanetary dust⁷. While infrared molecular spectroscopy has matured over the past half century, FT-IR microspectrometry is in its embryonic state, but is growing rapidly. FT-IR microspectrometry has been used to study polymeric materials⁸, yielding new insight into the distribution of phases in commercial plastics. Sections of brain tissue from Alzheimer's patients have been analysed⁹ to characterize the silicon-aluminium complex seen to accumulate in these brain cells. Extensive applications to electronic materials have been reported, from the analysis of defects in single crystals¹⁰ to the analysis of failed circuit boards¹¹. Uniting the microscope and the spectrometer gives the researcher a new tool with which to discover the properties of materials.

The success of FT-IR microspectrometry is a result of the discriminating power of the visual image combined with the analytical power of FT-IR spectroscopy—a case where the sum of the components is certainly much greater than that of the individual component parts. □

Joan Kwiatkoski and John Reffner are at Spectra-Tech, Inc., 652 Glenbrook Road, PO Box 2190G, Stamford, Connecticut 06906, USA. For more information, fill in reader service number 100.

- Coates, V.J., Offner, A. & Siegler, E.H. Jr *J. opt. Soc. Am.* **43**, 984 (1953).
- Steel, W.H. *Aust. J. Sci. Res.* **4**, 1 (1951).
- Barber, R., Cole, A.R.H. & Thompson, H.W. *Nature* **163**, 198 (1949).
- Gore, R.C. *Science* **110**, 710 (1949).
- Reffner, J.A., Coates, J.P. & Messerschmidt, R.M. *Am. Lab.* **19**, 4 (1987).
- Cook, R. & Paterson, M.D. *Forensic Sci. Inst.* **12**, 3 (1978).
- Fraudorf, P., Patel, R.I. & Freeman, J.J. *Icarus* **47**, 3 (1981).
- Harthcock, M.A., Lentz, L.A., Davis, B.L. & Krishnan, K. *Appl. Spectrosc.* **40**, 7 (1986).
- Hill, S.L., Krishnan, K., Pulsinelli, J. & Sutcliffe, J. *Pittsburgh Conference and Exposition on Analytical Chemistry and Applied Spectroscopy*, Abstr. No. 356 (1987).
- Ogawa, T. *Jap. J. Appl. Phys. (Part 2)* **25**, 11 (1986).
- Lang, P. & Katon, J.E. *Microbeam Anal.* **21**, 47–49 (1986).

ACS goes to New Orleans

Next week, 10,000 chemists will flock to New Orleans, Louisiana, for the American Chemical Society's annual meeting. Here are some of the products they will see in the over 400 exhibit booths.

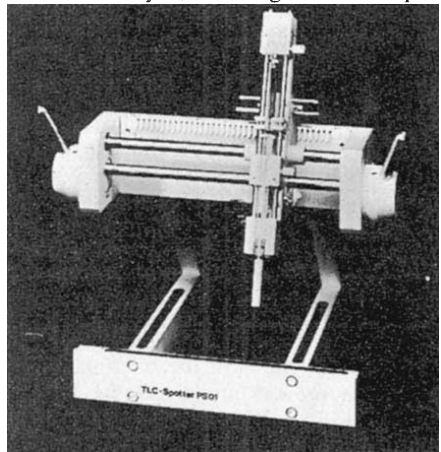
THE melting point of an unknown substance is one clue to its identity. Mettler has a **melting point instrument**, the model FP62, that determines melting points automatically (*Reader Service No. 101*). The instrument uses an optical detection



Mettler's instrument for melting points.

system that Mettler says provides corrected, true temperatures with high reproducibility. Heating rates can be selected between 0.1 and 10°C min⁻¹ in steps of 0.1°C. The temperature range extends from room temperature to 300°C. Results can be logged with a Mettler GA44 printer for a detailed printout of the results, or the melting curve can be recorded with an Epson-compatible printer. Mettler is exhibiting in booths 306 and 308.

Whatman will have information on its **semi-automatic TLC spotter** in booths 812 and 814 (*Reader Service No. 102*). The PSO1 model spotter is a precision contact spotter that dispenses the sample directly on the TLC layer, ensuring consistent spot

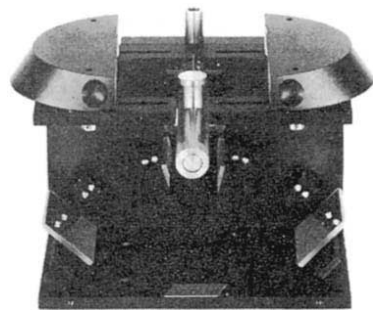


TLC spotting on a rail.

diameter. Whatman says spot size is kept to a minimum by the spotter's sensitive pulse motor. After the PSO1 automatically dispenses samples, it advances to the next pre-shot spotting position. Six 5 mm templates are available, that can direct up to 35 applications on a 20 × 20 cm TLC plate. The PSO1 costs £1,351 (UK).

The DPR-42 **liquid dipper ATR sampling system** from Analect Instruments, in booths 819 and 821, offers chemists the ability to do FT-IR sampling of stationary or flowing liquids rapidly and with little or no sample preparation (*Reader Service No. 103*). The system is designed to allow total isolation of sample and optics, allowing accurate measurements to be made without optics contamination by the sample or its vapour. The system comes in a fixed-position liquid dipper configuration, with a sealed liquid flow cell for the analysis of flowing liquids, or as an articulated dipper which can be rotated around the axis of the incoming IR beam, allowing the assembly to be tilted out of the way when changing samples. The stationary dipper costs \$3,950 (US), and the articulated dipper, \$4,750 (US).

Spectra-Tech, in booth 606, says it has an **FT-IR accessory** which can be used in



A close-up of Spectra-Tech's accessory.

diffuse reflectance infrared Fourier transform spectroscopy. Called the Collector, the accessory can be used in either specular or diffuse/specular mode (*Reader Service No. 104*). The collector features a thin metal blade positioned perpendicular to the surface of the sample cup that blocks energy that is reflected without penetrating the sample. Only diffusely reflected energy that has penetrated the sample can reach the collection mirror. Spectra-Tech says Reststrahlen bands are significantly reduced, and as a result, spectra are easier to interpret. Sample changing is speeded up by the Collector's slide-out sample holder that has a micrometer that allows

the sample height to be adjusted for maximum throughput of energy. The £2,775 (UK) collector comes with micro- and macro-sampling cups.

Dedicated detectors

A free brochure on Gilson Medical Electronics' model 112 **UV-visible detector** is now available, that explains its adaptability for a wide range of uses (*Reader Service No. 105*). The detector can be used for life science and industrial applications, using preparative and analytical HPLC systems, as well as gravity flow and low pressure chromatography. Its lamp/filter assemblies can be interchanged to study compounds from peptides and amino acids to haemoglobins, flavoproteins and commercial dyes. The model 112 offers detection at wavelengths of 214, 229, 254 and 280 nm, and in the range of 300–600 nm. Flow cells are available to accommodate sample sizes from micro-bore to preparative-scale. Gilson is exhibiting in booths 330 and 332.

Hewlett-Packard, in booths 412–420, is offering a \$5,000 discount from the US list



Diode-array detection on the cheap.

price of the model 1040M **diode-array detection system**, valid in the US until the end of the year (*Reader Service No. 106*). The discount brings the price of the system, including detector, HPLC ChemStation with colour monitor, HP colour-view software, 20 MByte Winchester disk drive, and printer, down to \$21,765 (US). HP says the Colourview software provides a comprehensive display of UV-visible absorbance data, and contains the latest developments for the ChemStation: UV-visible libraries, annotation and macro programming to customize operations.

All about chromatography

Scientific Systems, Inc. has an HPLC system that is suited for **instructional use** (*Reader Service No. 107*). The model 510 isocratic HPLC system features a rapid-refill pump and patented pulse damper for smooth flow. Its 3XL injector allows selection of three different internal sample loops, and settable high/low pressure shutoffs protect the system from damage. A prime/purge valve makes it easy to change mobile phases. Scientific Systems charges \$5,100 (US) for its system, and will be showing it off in booth 428.

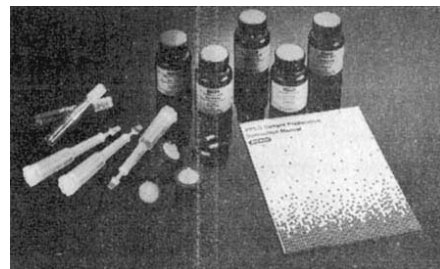
Perkin-Elmer, in booths 524–529, is offering an automatic **headspace sampling accessory** for its 8000 series gas chromatographs (*Reader Service No. 108*). The \$18,250 (US) model HS-101 uses the pneumatic sample transfer technique, eliminating the need for gas syringes and gas sampling loops. Perkin-Elmer says it is especially useful for the determination of residual solvents or monomers in polymers, haloforms in water and wastewater, blood alcohol, and volatiles in foodstuffs — but it can also be used for the identification of bacterial strains. Capillary columns may be passed through the heatable transfer tube to the sample needle, allowing splitless injection directly onto the capillary column. The capillary columns can also be used without cryofocusing. All samples of a series are thermostatted for the same amount of time, improving reproducibility of results and the analysis of thermolabile substances.

The System Two **supercritical fluid chromatography** system is the subject of a new product bulletin from Brownlee Labs (*Reader Service No. 109*). The System Two gives users complete control over both instrument operation and data handling. Pressure or density programming is selected by the user, and independent control of the dual syringe pump system allows special applications such as “doped” supercritical fluids or composition changes to be studied. The system has an uninterrupted, pulseless flow that eliminates the need to stop for refill. Standard chromatography software and software for specialized applications such as simulated distillation and hydrocarbon group separations are available. Brownlee will be displaying the System Two in booths 628–632.

Chromatofocusing is a simple, high resolution chromatographic method that separates proteins according to their isoelectric points in a self-generated gradient. Pharmacia has a new column, the Mono P HR 5/5, for **chromatofocusing** using its FPLC system (*Reader Service No. 110*). The 0.5 cm i.d. × 5 cm column is a weak ion exchanger. Proteins can be separated down to *pI* differences of 0.2, and the gel is stable between *pH* 2 and 12, allowing broad *pH* intervals during separation and effective column cleaning. The practical protein loading capacity is 5–10 mg ml⁻¹ media, and Pharmacia says the amount of non-specific adsorption is negligible and recovery of enzyme activity is normally 80 per cent. Pharmacia is in booth 343.

But first, sample prep.

Sample and reagent preparation is often the key to achieving sensitivity and efficiency in HPLC separations. Concentration steps, which allow detection of trace levels



Bio-Rad's spread for sample prep.

of compounds and removal of interfering compounds and particulates, are also necessary to assure maximum column life. Bio-Rad, in booth 703, has an **HPLC sample and reagent preparation kit** to help researchers select the optimum sample preparation method for their sample (*Reader Service No. 111*). The kit contains filters, disposable columns, and five ion exchange media, plus step-by-step instructions for each application.

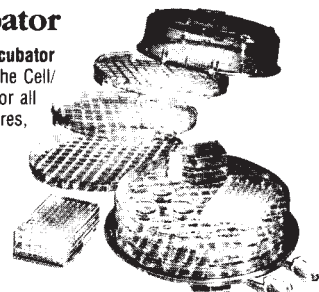
The spe-21 **column processing system** from J. T. Baker, in booths 713 and 715, lets researchers prepare up to 21 analytical samples simultaneously in 20 to 30 minutes (*Reader Service No. 112*). The columns are packed with Bakerbond bonded phases, and are arranged for convenient handling and uniform column flow. The system includes a vacuum gauge and controller, so the user can regulate the rate of extraction, wash and elution.

ADVERTISEMENTS

Versatile Tissue Culture Incubator

Modular Incubator Chamber (the Cell/Sphere™) for all tissue cultures, aerobic or anaerobic.

Low Cost



Eliminates evaporation loss and contamination. Easy to use — flush with gas mixture, seal, place at appropriate temperature. Stackable, see-through. Quantity discounts. **billups-rothenberg, inc., pob 977, Del Mar, CA USA 92014. Call 619-755-3309 or contact your Flow Laboratories representative.**

Reader Service No.174

Medwire Div. stocks a wide variety of **TEFLON COATED***

Platinum alloys, stainless steel, silver, multistrand and single strand available for prompt shipment. Write for descriptive brochure and price lists.

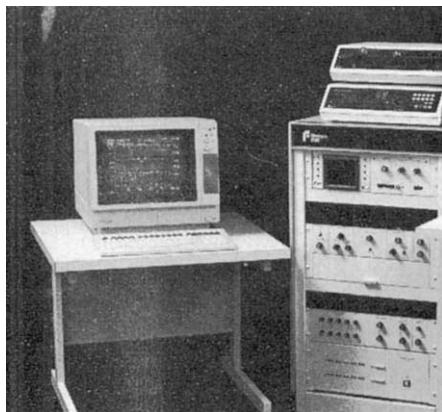


Medwire Div.

Affiliate of Sigmund Cohn Corp.
123 So. Columbus Ave.
Mt. Vernon, N.Y. 10553
914-664-5300

*Reg. T.M. of E.I. DuPont

Reader Service No.49



J.T. Baker's answer to multiple sample preparation needs.

The stainless steel sample collection needles are removable for cleaning, and the system's top is made of easy-to-clean Corian. The spe-21 system costs \$595 (US).

Varian, in booths 736–744, has a system that it says can slash sample preparation times by up to 80 per cent (*Reader Service No. 113*). The AASP Preperation is a pneumatically controlled **cassette processing station**, used in conjunction with the Varian Advanced Automated Sample Processor to automate sample preparation and injection into any HPLC. According to the company, the \$995 (US) Preperation provides increased reproducibility using controlled positive pressure. Precise flow restriction prevents pressure surges and cross-contamination of samples.

Mass spectrometry

Finnigan MAT, in booths 102 and 103, has a dedicated **thermospray liquid chromatograph/mass spectrometer system**, the TSP 46, that has three modes of soft ionization: conventional thermospray, filament and discharge chemical ionization (*Reader Service No. 114*). Using the filament for ionization, any mobile phase, including 100 per cent organic solvents such as those used in normal-phase HPLC, may be used. The discharge elec-

trode is used for chemical ionization in both polar and nonpolar solvents where the chromatographic requirements preclude the use of buffers. The \$170,000 (US) system can be used in conventional LC/MS applications, but was designed with the mass spectrometrists in mind, with features needed to identify complex molecules. The system's repeller electrode can be used with the discharge electrode to produce collision-induced dissociations (CID), that Finnigan MAT says when used with a single mass analyser provides CID/MS data comparable to those obtained with tandem MS/MS systems.

An overview of **mass spectrometry** is one of the papers in a 90-page booklet entitled *Focus on Mass Spectrometry*, published by Jeol (*Reader Service No. 115*). The booklet includes 19 recent papers on various aspects of mass spectrometry, including examples in the biomedical, pharmacokinetic and environmental fields. The stories behind the development of new MS equipment such as fast-atom bombardment (FAB)-MS, sparksources-MS, post-acceleration ion detectors, tandem MS-MS and TLC-FAB-MS are also included.

Designer molecules

ChemSmart is the name of the **chemical information software** package being demonstrated by the Institute of Scientific Information in booths 425 and 427 (*Reader Service No. 116*). ChemSmart comes with structure search and management software and a basic library of 250 organic compounds selected from the literature, and costs \$335 (US). Sixteen extra databases of 200 compounds each are available for \$100 (US), including anti-inflammatory agents, heterocycles, nucleosides/nucleotides, and herbicidal compounds. Each record consists of a graphic depiction of the compound's structure, and an electronic notebook with pertinent data. Searches can be performed for entire molecules, or fragments. The software

runs on an IBM PC/XT/AT or compatible with 256K RAM and a colour graphics board or enhanced graphics board.

Thioredoxin is a thiol protein with a molecular weight of approximately 12,000, that is used to unscramble proteins by acting as a disulphide interchange agent. Chemical Dynamics Corporation, in booth 1106, now offers **recombinant**



Unscrambles proteins with multiple cross-linkages by breaking disulphide bonds.

thioredoxin, produced in *Escherichia coli* (*Reader Service No. 117*). The company says thioredoxin treatment eliminates the disulphuration of cysteine and protein crosslinking experienced with current *in vitro* folding methods. The thioredoxin comes in 5 mg vials for \$75 (US) each.

FT-IR spectrometry

Mattson Instruments has a **benchtop FT-IR spectrometer** that operates at the push of a button (*Reader Service No. 118*). The £18,000 (UK) Polaris covers the spectral range $4,800\text{ cm}^{-1}$ to 400 cm^{-1} and produces a high-resolution infrared spectral analysis of a sample, says Mattson, within seconds of pressing two keys on the instrument's touch panel. Analyses can be performed without an external computer, with spectra produced by a digital plotter supplied with the instrument, or in conjunction with a Mattson ICON system or other IBM-compatible computer. The infrared beam enters and leaves the sampling compartment unfocused, so other techniques such as ATR, diffuse reflectance, specular reflectance, microtransmittance and photo-acoustic spectroscopy can also be performed. Mattson will be demonstrating the Polaris in booth 109.

From Nicolet, in booths 612–616, comes an **FT-IR spectrometer system** that has been designed to achieve the versatility and automated flexibility previously found only on larger mainframe instruments (*Reader Service No. 119*). Nicolet's system 740 has a frequency range from $10,000\text{ cm}^{-1}$ to 500 cm^{-1} , and costs £60,000 (UK). The instrument is equipped with Nicolet's data system 660, and FT-IR analysis software. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.



The TSP46 from Finnigan MAT is a versatile instrument for LC/MS.